R. Miller

JOURNAL OF

REGIONAL SCIENCE

Vol. 1 Summer 1958 No. 1

Walter Isard

Interregional Linear Programming: An Elementary Presentation and A General Model 1

Benjamin H. Stevens

An Interregional Linear Programming Model 60

John Q. Stewart and William Warntz

Physics of Population Distribution 99

Roy Gerard

Commuting and the Labor Market Area 124

PUBLISHED BY THE REGIONAL SCIENCE RESEARCH INSTITUTE

Vol. 1 Summer 1958 No. 1

JOURNAL OF REGIONAL SCIENCE

Edited by W. Alonso and W. Isard

The **Journal of Regional Science** is published two or more times a year by the Regional Science Research Institute, G.P.O. Box 8776, Philadelphia 1, Pa.

The **subscription rate** for libraries, business and other organizations, and individuals is $5.00 annually. Orders should be addressed to the Regional Science Research Institute. **Individuals only** may receive a subscription by joining the Regional Science Association and paying an extra charge for the **Journal**. Total yearly rate for subscription to the **Journal** and membership in the Association is currently $2.00. For further information about membership in the Association write to Regional Science Association, Wharton School, University of Pennsylvania, Philadelphia 4, Pa.

Back numbers and single issues can be obtained from the Institute while they are still in stock at a price of $2.50 per copy.

Manuscripts and other communications for the editors should be addressed to the Institute.

The Regional Science Research Institute is a non-profit corporation devoted to research and studies on the structure, function, and operation of regions from an economic, social, and political standpoint.

INTERREGIONAL LINEAR PROGRAMMING: AN ELEMENTARY PRESENTATION AND A GENERAL MODEL[1]

by Walter Isard

1. INTRODUCTION

This article has two objectives. One is to present in very elementary fashion the basic elements and procedures of interregional linear programming. This is done in the first sections. It is hoped that the discussion will make it possible for those unversed in mathematics and advanced social science analysis to grasp the fundamentals of this newly developing technique.

The second objective is to develop a rather pure, general model. It is not intended that this model as such be implemented. Rather it is the hope that the model will furnish a framework for the design of more specific interregional linear programs aimed at application to regional problems. This general model is evolved gradually from the elementary materials presented in the first sections.

Like interregional input-output and industrial complex techniques, interregional linear programming emphasizes general interdependence. In addition, it is an optimizing technique. It pertains to problems where the objective is to maximize or minimize some linear function, subject to certain linear inequalities. In such situations it purports to answer this kind of question: given a set of limited resources (which may include plant capacities, transportation and urban facilities, as well as mineral, labor

[1]The editors have thought it appropriate to devote a large part of this first issue of the **Journal** to interregional linear programming. This is a newly developing technique of analysis in the field of regional science and promises to make significant contributions. The first sections of this article are intended as an introduction to the subject, and readers already familiar with it may elect to skip them.

The author is deeply indebted to Benjamin H. Stevens for invaluable assistance in the development of his general model. Mr. Stevens' own, pioneering model, which is presented in the following article, "An Interregional Linear Programming Model," was well developed before the analysis of this paper was begun, and was generously made available for the author's use. In many ways, it has been heavily drawn upon.

The author is grateful to Resources for the Future, Inc. for a grant in support of his research.

and other natural and human endowments), given a technology in the form of a set of constant production coefficients, given a set of prices (except on the factors in limited supply), how program diverse production activities in order to maximize profits, social gains, total income, per capita income, employment, gross social product or some other magnitude? Or, how program diverse production activities in order to minimize transportation volumes, man-hours of work or some other magnitude, subject to the achievement of certain levels of output and consumption? In an interregional setting, the question can be still broader. For each of the several regions embraced in the study, there is given a set of limited resources, a technology (which may be the same from region to region, but which may lead to different production practices among regions because of different resource endowments), a set of prices on factors and commodities not in limited supply (where prices in the region are interrelated with prices in other regions through the existence, or possibility of trade), how program the diverse production and shipping activities of any given region in order to maximize income, employment or some other magnitude relating to the region? Or, more broadly, how program the diverse production and shipping activities of the several regions in order to maximize income, employment or some other magnitude of the interregional system (or nation, if the several regions do comprise a nation)?

As yet, there has not been a general linear programming model which has been both designed and profitably applied to a regional or interregional question of such broad scope as those posed in the above paragraphs. Yet there is considerable promise in this direction.[2] Some of the major difficulties encountered along the way will become evident in the discussion of the following sections. We begin at the most elementary level.[3]

---

[2]A series of significant steps have been taken and have already culminated in several important, though restricted, studies. E.g. see J.M. Henderson, "A Short-Run Model for the Coal Industry" and "Efficiency and Pricing in the Coal Industry," **Review of Economics and Statistics**, Vol. 37, Nov. 1955, pp. 336-346, and Vol. 38, Feb. 1956, pp. 50-60, respectively.

[3]Some general references on linear programming are: R.G.D. Allen, **Mathematical Economics**, MacMillan, London, 1956, chaps. 16-19; A. Charnes, W.W. Cooper and A. Henderson, **An Introduction to Linear Programming**, John Wiley & Sons, New York, 1953; J. Chipman, "Linear Programming" and "Computational Problems in Linear Programming," **Review of Economics and Statistics**, Vol. 35, May 1953, pp. 101-117, and Nov. 1953, pp. 342-349, respectively; R. Dorfman, P.A. Samuelson and R.M. Solow, **Linear Programming and Economic Analysis**, McGraw-Hill, 1958; and T.C. Koopmans (ed.), **Activity Analysis of Production and Allocation**, John Wiley and Sons, New York, 1951.

Among others, the following papers deal in one way or another with regional and locational aspects of linear programming: M. Beckman and T. Marschak, "An Activity Analysis Approach to Location Theory," **Kyklos**, Fasc. 2, 1955; Edward B. Berman, "A Model for Maximizing a Vector of Final Demand Deliveries under Regional Production and Transportation Network Constraints," (mimeo, 1957); W.L. Garrison and D.F. Marble, "The Analysis of Highway Networks: A Linear Programming Formulation," (mimeo, 1958); T.A. Goldman, "Efficient Transportation and Industrial Location," **Papers and Proceedings of the Regional Science Association**, Vol. 4, 1958; J.M. Henderson, **op. cit.**, and "The Utilization of Agricultural Land: A Regional Approach," **Papers and Proceedings of the Regional Science Association**, Vol. 3, 1957, pp. 99-114; T.C. Koopmans, **op. cit.**, chaps. 14 and 23; L. Lefeber, "General Equilibrium Analysis of Production, Transportation and the Choice of

2. A SIMPLE LINEAR PROGRAMMING PROBLEM: A GRAPHIC SOLUTION

To illustrate the use of the technique on a specific, over-simplified level, consider a hypothetical isolated region desiring to realize a maximum of new income from productive activities, but having available for such purposes only a limited quantity of each of four resources: water, land, labor, and capital, the last being free for investment. Two economic activities have been identified as "profitable". To generate one dollar of new income, each activity requires a set of inputs of each of these resources as listed in Table 1. (In linear programming, we must generally assume, as we do in this problem, constant production coefficients and fixed prices on all commodities but resources in limited supply.)

TABLE 1. Resource Requirements per Dollar New Income.

Required Units of:	Activities #1	#2
Water	0.5	0.6
Land	0.2	0.15
Labor	0.4	0.2
Capital	3.0	2.0

If we define the unit level of operation of each activity as that level which generates one dollar of new income, the data of Table 1 refer to unit levels of operations. We now desire to find the most desirable combination of levels (i.e. multiples of these unit levels) at which to operate these activities. Since these levels are unknowns, they may be designated X_1 and X_2, respectively. To work with, the region has only 6MM units of water, 1.8MM units of land, 3MM units of labor and 24MM units of capital.

The problem can be solved rather easily with the use of a graph. Along the vertical axis of Figure 1 we measure level of activity #1; along the horizontal level of activity #2. Next we construct resource limitation lines. The water limitation line NU indicates the various combinations of

---

Industrial Location," Papers and Proceedings of the Regional Science Association, Vol. 4, 1958, and Allocation in Space, North-Holland Publishing Co., Amsterdam (forthcoming); F.T. Moore, "Regional Economic Reaction Paths," American Economic Review, Vol. 45, May 1955, pp. 133-148; L.N. Moses, "An Input-Output, Linear Programming Approach to Interregional Analysis," Report, 1956-57, Harvard Economic Research Project; E.W. Orr, "A Synthesis of Theories of Location, of Transport Rates, and of Spatial Price Equilibrium," Papers and Proceedings of the Regional Science Association, Vol. 3, 1957, pp. 61-73; P.A. Samuelson, "Spatial Price Equilibrium and Linear Programming," American Economic Review, Vol. 42, June 1952, pp. 152-162; B.H. Stevens, op. cit.; and T. Vietorisz, Regional Programming Models and the Case Study of a Refinery-Petrochemical-Synthetic Fiber Industrial Complex for Puerto Rico, Ph.D. dissertation, M.I.T., 1956, Part I.

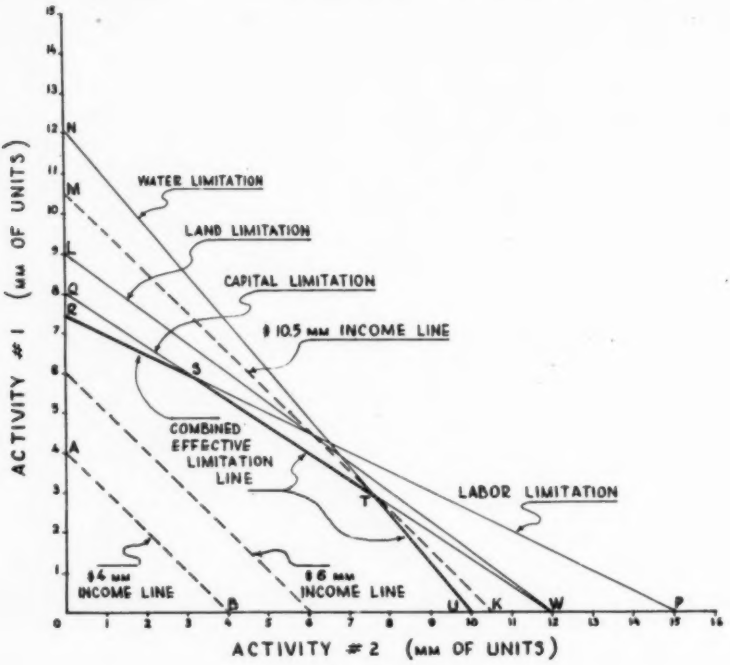


FIGURE 1.

levels of activities #1 and #2, whose water requirements do not exceed the **6MM** units available for consumption. At one extreme (given by point **N**) operation of activity #1 at a level of **12MM** units could be achieved provided activity #2 were carried on at zero level. At the other extreme (given by point **U**) operation of activity #2 at a level of **10MM** units could be achieved, provided activity #1 were carried on at zero level. Along the line and in between points **N** and **U** are the various combinations of levels of the two activities which just exhaust the available **6MM** units of water. Below and to the left of line **NU** are an infinite number of points, each of which however represent combinations of levels which require less than **6MM** units of water. Since for any given point below and to the left of line **NU** there can always be found a point on line **NU** which corresponds to greater levels of both activities, all points in the positive quadrant bounded by **NU** (but not including **NU**) may be said to represent **inefficient** combinations **so far as the water limitation is pertinent.**

However, if one moves up along the vertical axis (corresponding to zero level of activity #2), he finds that well before the water limitation on the level of activity #1 takes effect, the limitations on labor, capital and land have become effective. At point **R**, for example, the labor limitation becomes operative. At this extreme point, given only **3MM** units of labor and the requirement of 0.4 units of labor at unit level, the maximum

level at which activity #1 can be operated is 7.5MM units. Thus it becomes necessary to consider other resource limitations. To do so, we have constructed the labor limitation line RP, the capital limitation line QW, and the land limitation line LW. These lines are constructed on the same basis as the water limitation line. For example, point P at the other extreme of the labor limitation line indicates that activity #2 can be carried on at a maximum level of 15MM units provided activity #1 is pursued at zero level. And points along line RP indicate all combinations of levels of the two activities whose requirements of labor just equal the available 3MM units. Points below and to the left of the labor limitation line RP are inefficient, so far as the labor limitation is pertinent.

In this problem the implications of all resource limitations must be considered. It is clear that any combination of levels represented by a point in the area above the stretch RS (e.g. the combination of 7.7MM units of activity #1 and 0.1MM units of activity #2) is unattainable. Although such a combination may consume amounts of capital, land, and water which do not exceed the available quantities of these resources, it does require labor in excess of available supply. Hence the labor limitation line, in particular stretch RS, is binding in this area of the positive quadrant. Passing to the right of point S -- or more strictly to the right of a line drawn from the origin through S -- another resource limitation line, namely, that representing capital, becomes binding. In this area, any point representing a combination of activity levels which is feasible from the standpoint of available capital is also feasible from the standpoint of other available resources. On the other hand, any point representing a combination which is feasible from the standpoint of these other resources is not necessarily feasible from the standpoint of capital. For example, combinations represented by points directly above the stretch ST require a capital supply in excess of that available, although these combinations do not exceed other resource limitations. Hence, the capital limitation line, in particular stretch ST, is effective in this area.

Going still further to the right beyond T -- more strictly to the area to the right of the ray from the origin through T -- another resource limitation becomes binding, namely water. Water is the first resource whose available supply is fully utilized by combinations of levels of activities #1 and #2 which can be represented by points in this area. In particular, the stretch TU of the water limitation line becomes effective.

In brief, then, the segmented line RSTU delineates a closed set of points in the positive quadrant which represent all combinations of activity levels which are feasible from the standpoint of each and every resource limitation. These are attainable points. Combinations represented by points in the positive quadrant but outside this closed set require at least one resource in excess of available supply. These combinations are not feasible; the corresponding points are unattainable. Further, since it can be demonstrated that any point in this closed area not on RSTU is inferior to some point on the line, in the sense that the latter point corresponds to greater achievable levels for both activities, we can state that all efficient points of the feasible set of points lie on the segmented line RSTU.[4]

[4] In linear programming parlance, the segmented line RSTU is designated the "efficiency frontier," or in mathematical parlance the "convex hull" of attainable points.

Having eliminated "inefficient" combinations of activity levels, we now seek to determine that efficient combination which will maximize new income, i.e. the optimal combination. To do this is relatively simple since for each activity we have already defined the unit level as that level which generates $1 of new income. Following standard procedures in economic analysis, we may construct "iso-income" lines. For example, point A in Figure 1 represents a combination of a 4MM unit level of activity #1 and a zero unit level of activity #2; therefore, it corresponds to $4MM of new income. Point B represents a combination of a zero unit level of activity #1 and a 4MM unit level of activity #2; it also corresponds to $4MM of new income. Likewise, combinations of levels represented by all points lying along a straight line from A to B yield $4MM of income. Thus we have a $4MM income line (dashed in Figure 1). Similarly, we can obtain a $6MM income line (dashed in Figure 1) and other lines each representing a locus of points which represent combinations yielding the same total income, and each with a slope of -1 when the scales along the vertical and horizontal axes are identical. Since the "iso-income" lines increase in value as one moves farther and farther from the origin, we wish to identify the highest "iso-income" line on which any of our "efficient" points (i.e. points on RSTU) lie. Such a line is the $10.5MM income line upon which point T lies. Thus point T represents the combination of activity levels (namely, 3MM units for #1 and 7.5MM units for #2) which maximizes new income. It is the optimal solution, which is therefore the solution to our simple problem.

3. A SIMPLEX-TYPE COMPUTATION

The above type of graphic solution is simple and effective when only two activities are considered. Unfortunately, the graphic solution becomes complex when a third activity is introduced into the problem. Three dimensions must then be used, the level of the third activity being measured along the third dimension. And when we introduce a fourth activity, a fifth, a sixth...and finally an nth activity into the problem, a direct graphic solution is not possible. It is at this point that other types of solutions must be sought to the linear programming problem. It also becomes desirable at this point to state the problem in more formal, mathematical terms.

To begin, suppose n = 3, i.e. that there are three activities to be considered. Associated with each activity are a set of data indicating inputs of each resource required in order for that activity to operate at unit level and thus generate one dollar of new income. We represent these data by a set of coefficients as follows:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix}$$

If the data of Table 1 are still relevant for the first two activities; then a_{21} which represents the requirement of resource #2 (namely, land) per unit

level of **activity #1** is 0.2; a_{12} which represents the requirement of **resource #1** (namely, water) per unit level of **activity #2** is 0.6; a_{32} which represents the requirement of **resource #3** (namely, labor) per unit level of **activity #2** is 0.2; etc. Letting X_1, X_2, and X_3 represent the levels of activities #1, #2, and #3, respectively, which are to be determined and which are restricted to non-negative values we wish to maximize income which is equal to **$1.00** x Z where

(1) $$Z = X_1 + X_2 + X_3;$$

or more specifically, to maximize Z (i.e. the number of dollar generating units) subject to the conditions that the sum of the requirement of any one resource by all activities not exceed the supply of that resource. If we denote R_1, R_2, R_3, and R_4 as the available supplies of water, land, labor and capital, respectively (which in our problem are **6MM, 1.8MM, 3MM** and **24MM** units, respectively), and if we note, for example, that $a_{12}X_2$ represents the requirement of the first resource by the second activity [since this involves multiplying the requirements of resource #1 per unit level of activity #2 (i.e. a_{12}), by the level of activity #2 (i.e. X_2)], the above conditions can be written:

(2)
$$\begin{aligned} a_{11}X_1 + a_{12}X_2 + a_{13}X_3 &\leq R_1 \\ a_{21}X_1 + a_{22}X_2 + a_{23}X_3 &\leq R_2 \\ a_{31}X_1 + a_{32}X_2 + a_{33}X_3 &\leq R_3 \\ a_{41}X_1 + a_{42}X_2 + a_{43}X_3 &\leq R_4 \end{aligned}$$

The first of these inequalities, for example, states that the requirement of resource #1 by **activity #1 plus** the requirement of resource #1 by activity #2 **plus** the requirement of resource #1 by activity #3 is less than or equal to the available supply of resource #1.

Note that our problem conforms to the general framework of linear programming. We are maximizing Z, which is a **linear** function of non-negative variables (X_1, X_2, and X_3) subject to four restraints, each one of which is a **linear** inequality.

Once the problem has been appropriately formulated, a computational procedure must be adopted. There are several which are possible,[5] the one in most general use being the simplex method. The simplex method is one of iteration, and typically involves in a many resource-many activity problem an extensive set of computations for which a high-speed computing machine is utilized.

In order to illustrate the simplex method in a manageable way, we shall confine our problem to only two limited resources (say water and land) and three activities. We therefore wish to maximize income, i.e. $1 x Z where

(3) $$Z = X_1 + X_2 + X_3$$

[5] See Koopmans, **op. cit.**, and Dorfman, Samuelson and Solow, **op. cit.**

subject to:

$$(4) \qquad \begin{aligned} a_{11}X_1 + a_{12}X_2 + a_{13}X_3 &\leq R_1 \\ a_{21}X_1 + a_{22}X_2 + a_{23}X_3 &\leq R_2 \end{aligned}$$

and where $X_1 \geq 0$; $X_2 \geq 0$; $X_3 \geq 0$. The first step is to change the inequalities (4) into a set of equations by introducing slack variables. Each slack variable refers to the level at which we operate a corresponding disposal activity, where a disposal activity can dispose of a resource at zero cost and does not give rise to any additional income. For the sake of convenience, the unit level of a disposal activity may be defined as that level which disposes of a unit of the relevant resource. In our problem we may therefore inject two new activities: a fourth activity which disposes of water and whose level is X_4, and a fifth activity which disposes of idle land and whose level is X_5. Thus, for non-negative values of X_1, X_2, X_3, X_4 and X_5 we wish to maximize income, namely Y, where:

$$(5) \qquad Y = c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 + c_5X_5$$

in which c_1, c_2 and c_3 represent the new income from operating at unit level activities #1, #2, and #3, respectively, and which in our problem have the value of \$1; and c_4 and c_5 represent new income from operating at unit level the disposal activities #4 and #5, respectively, and which by definition have the value of \$0.00. In maximizing Y, we are subject to the constraints:

$$(6) \qquad \begin{aligned} a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{14}X_4 + a_{15}X_5 &= R_1 \\ a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + a_{24}X_4 + a_{25}X_5 &= R_2 \end{aligned}$$

The first three coefficients in each of equations (6) are the corresponding coefficients of the resource input matrix represented above. The coefficient a_{14} is unity indicating the using up of one unit of water per unit level of operation of the water disposal activity while the coefficient a_{24} is zero indicating that no land inputs are required by the water disposal activity. For analogous reasons, a_{15} is zero and a_{25} is unity. Thus, for example the first of equations (6) states that the requirement of water by the first activity plus the requirement of water by the second activity, plus the requirement of water by the third activity plus the requirement of water by the water disposal activity plus the requirement of water by the land disposal activity (which requirement is zero) equals the total supply of water available.[6]

Next we select a set of values for the X's which satisfy equations (6). Note that we have only two such equations. If we so desired, we could con-

[6] If we refer to the graphic solution of the above problem which involved two activities and four resources, we observe that the available supply both of water and capital were fully employed, but that the requirements of land fell short of available supply by 1.275MM units. If we were to introduce into that problem a land disposal activity which used up one unit of land per unit level of operations, and which operated at a level of 1.275MM units, the total requirements of land would equal total supply available, i.e. $a_{21}X_1 + a_{22}X_2 + a_{25}X_5$ would equal R_2.

sider any two of the five variables involved by setting the other three variables equal to zero, in essence eliminating their corresponding activities from the problem. We would then have two equations in two unknowns. Provided these equations (the initial constraints) are independent and are not inconsistent we could solve for these two unknowns; their values would represent the levels at which corresponding activities are to be operated. For example, if we take as the relevant coefficients those indicated in Table

TABLE 2. Resource Requirements per Unit Level of Activity.

	#1	#2	#3	#4	#5
1. Water	0.5	0.6	0.7	1.0	0.0
2. Land	0.2	0.15	0.1	0.0	1.0

2,[7] and if we retain the values of 6MM and 1.8MM units for R_1 and R_2, respectively, we have the following two equations:

$$(7) \qquad \begin{aligned} 0.5X_1 + 0.6X_2 + 0.7X_3 + 1.0X_4 + 0.0X_5 &= 6MM \\ 0.2X_1 + 0.15X_2 + 0.1X_3 + 0.0X_4 + 1.0X_5 &= 1.8MM \end{aligned}$$

If we now set $X_1 = X_2 = X_3 = 0$, we obtain:

$$(8) \qquad \begin{aligned} 1.0X_4 + 0.0X_5 &= 6MM \\ 0.0X_4 + 1.0X_5 &= 1.8MM. \end{aligned}$$

Immediately apparent is the solution to equations (8), which gives $X_4 = 6MM$; and $X_5 = 1.8MM$.

In the same manner, we could obtain solutions for every other pair of variables provided we confront consistent and independent equations.[8] Further, there are still many other solutions possible. There are many with positive values for three variables, and zero for the other two (e.g. $X_1 = 1MM$; $X_2 = X_3 = 0$; $X_4 = 5.5MM$; and $X_5 = 1.6MM$). There are also many with positive values for four variables, and zero for the other (e.g. $X_1 = 2MM$; $X_2 = 3MM$; $X_3 = 0$; $X_4 = 3.2MM$; $X_5 = 0.95MM$). Finally, there are

[7] In this table of coefficients, a_{11}, a_{12}, a_{21}, and a_{22} (0.5; 0.6; 0.2; and 0.15) are taken from Table 1. The coefficients a_{13} and a_{23} (0.7 and 0.1) are new and indicate respectively the requirements of water and land per unit level of the third activity. The coefficients a_{14}, a_{15}, a_{24}, and a_{25} (1.0; 0.0; 0.0; and 1.0) are in keeping with the statements already made in the text.

[8] Some of these might involve a negative value for one of the variables, and hence would be discarded since we are limiting solutions to those which involve non-negative values for all variables. For example, the solution for equations (7) where $X_2 = X_3 = X_4 = 0$ is $X_1 = 12MM$ and $X_5 = -0.6MM$; this solution is excluded.

many with positive values for all five variables.

In our problem, however, we are interested in finding that solution (or that set of solutions each of which is associated with the same income) which yields maximum income. Since we are dealing with linear equations we can use a simple logical process to reduce by manyfold the number of necessary computations. Take a solution which involves more than two variables, say the one where $X_1 = 4.5MM$; $X_2 = X_3 = 0$; $X_4 = 3.75MM$; and $X_5 = 0.9MM$. This solution may be termed a three-variable solution since only three variables have values greater than zero, the others being zero. Further, this three-variable solution can be viewed as a weighted average of two other solutions, each involving only two variables, where the weights add to unity. In this specific case the first of these other two-variable solutions is:

$$X_1 = 9MM; \text{ and } X_4 = 1.5MM$$

and the second is:

$$X_4 = 6MM; \text{ and } X_5 = 1.8MM,$$

a two variable solution which we already noted. If we multiply the first solution by 0.5 and the second solution by 0.5 we obtain the above solution in three variables.

Now suppose we calculate according to equation (5) the income generated by each of these two two-variable solutions. Suppose the first has a higher income than the second. Then, since the solution in three variables is an average of the two two-variable solutions, it must follow that the income associated with the three-variable solution must be less than the income associated with the better of the two-variable solutions. Thus, the better of the two-variable solutions is to be preferred to the three-variable solution.

However, it might develop that both of the two-variable solutions are equally good, i.e. yield the same income. Then the three-variable solution being an average of these two solutions yields the identical income. We therefore conclude that no three-variable solution can have an income associated with it which is greater than the income associated with the best of two-variable solutions, where the best may refer to a **single** solution or to **several** solutions if more than one yield the same maximum income.

In the same manner, it can be shown that any four-variable solution can be considered a weighted average of a set of two variable solutions, where the weights add to unity. Since such a solution is an average, the income associated with it cannot be any higher than the income associated with the best (one, or several) of the two-variable solutions. At most it can only be equal to the income associated with the best of the two-variable solutions. Similarly, any five-variable solution can be expressed as a weighted average of two-variable solutions, and hence cannot have an income associated with it which is any greater than the income of the best (one, or several) two-variable solutions.

As a result of this simple logic we can say that if from among the two-variable solutions we find that one (or the several) which yields the

highest income we know that no other solution involving more than two variables can yield a higher income. We therefore short-cut the computation procedure for our problem by paying attention solely to two-variable solutions. The simplex method is one of several methods which aims at identifying the best (one or several) of these two-variable solutions.

For our problem, the procedure of the simplex method is as follows. We pick a two-variable solution which is usually called a **basic** solution.[9] For example, take our solution where X_4 = 6MM and X_5 = 1.8MM. Associated with this solution is an income which is calculated by substituting these values into equation (6). The resulting income is zero. Since we suspect that we can find a higher income solution, let us consider another two-variable solution. But which one? Several are possible. In the simplex method -- which aims to examine systematically the various two-variable solutions in a relatively efficient manner -- the next step is to substitute for one of the activities in the first basic solution another activity which was formerly excluded (i.e. operated at zero level). To determine which activity should be excluded and which one should take its place, we may develop a table such as Table 3. In Table 3 the first column indicates the activities at positive levels in the initial basic solution. The second column indicates the levels at which each of these is operated. The third column indicates income per unit level of each of these activities.

Thus far, we record data already noted. In columns 4-6, we present new data. These new data indicate the **combination** of activities at positive levels in the initial basic solution, namely activities #4 and #5, which are technically equivalent to a unit level of each one of the three "excluded" activities, i.e. activities which are not included in the first basic solution. For example, the technical requirements of activity #1 at unit level are 0.5 units of water and 0.2 units of land, as given by the coefficients of Table 2. If we reduce activity #4 by 0.5 units we will save 0.5 units of water, and none of land (since by Table 2 a unit level of activity #4 requires 1.0 units of water and 0.0 units of land). If we reduce activity #5 by 0.2 units, we will reduce requirements of water by 0.0 units and of land by 0.2 units (since by Table 2 a unit level of activity #5 requires 0.0 units of water and 1.0 units of land). Thus, our total requirements of water and land remain the same if we increase activity #1 by 1.0 units and decrease activity #4 by 0.5 units and activity #5 by 0.2 units. Thus, a unit level of activity #1 is technically equivalent to the combination of 0.5 units of activity #4 plus 0.2 units of activity #5. This combination is recorded in column 4 of Table 3.

In like manner we obtain the data for columns 5 and 6. As indicated in column 5, a unit level of activity #2 requires the same quantities of water and land as the combination of 0.6 units of activity #4 and 0.15 units of activity #5; it is therefore technically equivalent (substitutable) for this

[9] A basic solution is a solution which involves no more variables than there are constraints. When the values of these variables are non-negative the basic solution is also feasible. A non-basic solution is a solution which involves more variables than there are constraints. It is also feasible when the values of these variables are non-negative.

Since one of the possible optimal solutions will be a basic solution, one can always obtain an optimal solution which involves no more positive level activities than there are constraints.

TABLE 3.

1st Basic Solution: Activities at Positive Levels	Level at which Operated	Income Per Unit Level	Combination of Activities at Positive Levels in Basic Solution Technically Equivalent to (for which can be substituted) a unit level of:		
			Activity #1	Activity #2	Activity #3
(1)	(2)	(3)	(4)	(5)	(6)
#4	6MM	$0.00	0.5	0.6	0.7
#5	1.8MM	0.00	0.2	0.15	0.1
A. Income Generated by Above Combinations of Activities #4 and #5			$0.00	$0.00	$0.00
B. New Income Generated by Unit Level of			1.00	1.00	1.00
C. Difference (B-A) = Net Gain From Substituting for Above Combination a unit level of			1.00	1.00	1.00

An Excluded Activity Associated with the Maximum Difference: #1
Number of units of Activity #1 which can be substituted into the problem until the level of { activity #4 falls to zero: 6MM/0.5 = 12MM; activity #5 falls to zero: 1.8MM/0.2 = 9MM

combination. Similarly, as column 6 notes, a unit level of activity #3 is technically equivalent to the combination of 0.7 units of activity #4 and 0.1 units of activity #5.

The next step is to determine whether it is profitable to substitute in our solution a unit of any "excluded" activity for its technically equivalent combination of activities #4 and #5. We therefore ask: (1) what will be the direct gain in income from operating an excluded activity at unit level? (2) what will be the corresponding loss in income by curtailing the operation of activities #4 and #5 in order to release the necessary resources? and (3) is the gain greater or smaller than the loss? The answers to these questions are given in rows A, B, and C of Table 3. To illustrate, take activity #1. To increase it from zero to unit level necessitates first the curtailment of activity #4 by 0.5 units which results in a zero loss of income since income from a unit level of operation of activity #4 (the water disposal activity) is zero; and second the reduction of activity #5 by 0.2

units which also results in a zero loss of income since income from a unit level of operation of activity #5 (the land disposal activity) is zero. All told the loss of income, i.e. the income generated by the above combination of activities #4 and #5, is zero. This figure is given in row **A**, column 4, Table 3. The new income generated by increasing activity #1 from zero level to unit level is $1.00, as defined above in connection with equation (5). This figure is noted in row **B**, column 4. The difference between the figures in rows **B** and **A** of column 4 is **$1.00** and represents the net gain in income from increasing activity #1 by a unit level while necessarily decreasing activities #4 and #5 by that combination of levels technically equivalent to a unit level of activity #1 in the use of resources. This difference is given in row **C**, column 4. Since this difference is positive, it is profitable to substitute activity #1 for its technically equivalent combination of activities #4 and #5.

But perhaps it is still more profitable to substitute still another "excluded" activity, say activity #2 or #3, for its equivalent combination of activities #4 and #5. To determine whether or not this is the case, we perform the same computation for each of these activities, the resulting data being recorded in Table 3. Looking at Table 3, in particular row **C**, we note that it is as equally profitable to substitute a unit level of activity #2 (or #3) for its equivalent combination of activities #4 and #5 as it is to substitute a unit level of activity #1 for its corresponding equivalent combination. To break the deadlock, we choose to substitute activity #1 for activities #4 and #5; i.e. our procedure is to choose to substitute for a combination of the included activities an "excluded" activity associated with the maximum difference (recorded in row **C**).

Since in our framework each activity operates at constant unit cost, if it is profitable to substitute one unit of activity #1 for an equivalent combination of activities #4 and #5, it is profitable to substitute many units for a like multiple of this combination. But the terms of reference in our problem state that no activity can be operated at a negative level. Therefore, all we can do is to continue to substitute unit levels of activity #1 for equivalent combinations of 0.5 units of activity #4 and 0.2 units of activity #5 until the overall level of one of these activities is reduced to zero. From the calculations at the bottom of Table 3, we note that activity #4 will not reach the zero level until we have substituted in 6MM/0.5 or 12MM units of activity #1; whereas activity #5 will not reach the zero level until we have substituted in 1.8MM/0.2 or 9MM units of activity #1. We therefore are allowed to substitute only 9MM units of activity #1 into the problem. When this occurs, the operation of activity #5 has been reduced to zero; the level of activity #5 has been eliminated as a variable.

Our next step is to consider the superior solution in which activities #1 and #4 are operated at positive levels. (We have demonstrated that a solution with activities #1, #4, and #5, as well as one with just activities #4 and #5, is inferior.) We return to equations (7). We set $X_2 = X_3 = X_5 = 0$. We solve and find that $X_1 = 9MM$ units and $X_4 = 1.5MM$.[10] We call this the second basic solution. We are now ready to set up a second table, Table 4, which evaluates the second basic solution.

[10]We could have derived this solution directly from Table 3 since we had already established the fact that X_1 should be 9MM units (the level at which one of the previously included activities first becomes zero).

TABLE 4.

2nd Basic Solution: Activities at Positive Levels	Level at which Operated	Income Per Unit Level	Combination of Activities at Positive Levels in Basic Solution Technically Equivalent to (for which can be substituted) a unit level of:	
			Activity #2	Activity #3
(1)	(2)	(3)	(4)	(5)
#1	9MM	$1.00	0.75	0.5
#4	1.5MM	0.00	0.225	0.45
A. Income Generated by Above Combination of Activities #1 and #4			$0.75	$0.50
B. New Income Generated by Unit Level of:			1.00	1.00
C. Difference (B-A) = Net Gain From Substituting for Above Combination a Unit level of:			0.25	0.50

Activity Associated with Maximum Positive Difference: #3
Number of units of Activity #3 which can be substituted into the problem until the level of { activity #1 falls to zero: 9MM/0.5 = 18MM; activity #4 falls to zero: 1.5MM/0.45 = 3.33MM }

In Table 4 the first column indicates the activities to be operated at positive levels. The second column indicates the levels at which they are operated. The third column indicates income per unit level of each of the included activities (as given in equation 5). The fourth and fifth columns pertain to activities #2 and #3 which are "excluded" from this second basic solution. (Note that there is no column for the excluded activity #5. This is so because it can be demonstrated -- as intuitively seems logical -- that once an activity has been eliminated in proceeding from one basic solution to a second, superior basic solution, it will not be one of the included activities in the one or more basic solutions which will be associated with maximum income. Therefore, we need not consider reintroducing activity #5 by substitution for an equivalent combination of "included" activities.)

We now investigate the profitability of substituting one of the excluded activities (#2 or #3) for one of the included ones (#1 or #4). The first two items in columns 4 and 5 indicate the combinations of activities #1 and #4 to which a unit level of activity #2 and #3 are respectively technically equivalent. Thus, for example, activity #2 is technically equivalent to

a combination of 0.75 units of activity #1 and 0.225 units of activity #4. That is, if we reduce activity #1 by 0.75 units and activity #4 by 0.225 units we will release just enough water and land to make possible a unit level of activity #2. If we do reduce activity #1 by 0.75 units, we lose $0.75 of income since activity #1 generates $1 income per unit level, and if we do reduce activity #4 by 0.225 units we lose zero income since activity #4 generates zero income per unit level. All told, if we reduce activity #1 by 0.75 units and activity #4 by 0.225 units we lose $0.75 income, which is recorded in row **A**, column 4, Table 4. But as we reduce these activities by the above amounts, we can increase activity #2 by one unit. This increase generates $1 of new income, since income per unit level of activity #2 is $1. This $1 is recorded in row **B**, column 4. The difference between the gain and the loss, namely, $0.25, is recorded in row **C**, column 4. Thus our total income would rise by $0.25 every time we substitute one unit of activity #2 for 0.75 units of activity #1 and 0.225 units of activity #4.

In similar manner we perform the same computations for activity #3 (and in the general case for any other "excluded" activity which has not previously been eliminated). The results of this computation are recorded in column 5 of Table 4.

As we look over row **C** of Table 4 we note that the maximum gain per unit level of any excluded activity is associated with activity #3. We therefore wish to substitute units of activity #3 for technically equivalent combinations of activities #1 and #4. We can proceed with this substitution until the level of either activity #1 or #4 falls to zero. There we must stop, since no activity is permitted to be operated at negative levels. From the bottom of Table 4 we note that the level of activity #1 falls to zero when 9MM/0.5 = 18MM units of activity #3 are substituted into the solution, and that the level of activity #4 falls to zero when 1.5MM/0.45 = 3.33MM units of activity #3 are substituted in. Thus, we substitute 3.33MM units of activity #3 into the solution at which point activity #4 falls to a level of zero. At that point we have a third basic solution consisting of the two activities #1 and #3. In equations (5) we let $X_2 = X_4 = X_5 = 0$ and solve. We find that $X_1 = 7.33$MM units and $X_3 = 3.33$MM units. This basic solution is superior to either of the first two basic solutions, or to any three or four variable solution consisting of activities #1 and #3 (which are included in this third basic solution) and activities #4 and #5 (which have been eliminated).

We still must ask: is there any other basic solution which is superior to the best one we have found thus far, namely, the third basic solution? To answer this question we once again carry through a set of computations. The results are presented in Table 5 which is constructed in the same manner as Tables 3 and 4. Since two of the three variables excluded from this third basic solution, namely, activities #4 and #5 have already been eliminated, we need a column for only one activity, namely, activity #2 which is the only activity to be considered for substitution into this third basic solution. However, we have added a column for activity #5, which was the first activity to be eliminated. We have done so in order to illustrate the unprofitability of considering excluded activities which have already been eliminated from previous basic solutions. If one considers the data pertaining to activity #5 in column 5, he notes that for every unit of activity #5 which might be substituted for its technically equivalent com-

TABLE 5.

3rd Basic Solution: Activities at Positive Levels	Level at which Operated	Income Per Unit Level	Combination of Activities at Positive Levels in Basic Solution Technically Equivalent to (for which can be substituted) a unit level of:	
			Activity #2	Activity #5
(1)	(2)	(3)	(4)	(5)
#1	7.33MM	$1.00	0.5	7.77
#3	3.33MM	1.00	0.5	-5.55
A. New Income Generated by Above Combination of Activities #1 and #3:			$1.00	2.22
B. New Income Generated by Unit Level of:			1.00	0.00
C. Difference (B-A) = Net Gain from Substituting for Above Combination a Unit Level of:			0.00	-2.22
Activity Associated with Maximum Positive Difference: None				

bination of activities #1 and #3, there is a loss of $2.22.

Observing the data in column 4 which pertains to activity #2, we note that the increase in income resulting from a unit increase in the level of activity #2 is just matched by the loss of income from the necessary reduction of 0.5 units in the levels of both activities #1 and #3. The net gain as recorded in row C of column 4 is zero. Therefore, it is not possible to increase total income by substituting into the third basic solution a unit of activity #2 (or any other activity) for its technically equivalent combination of activities #1 and #3. We have come to the end of our journey. We have found a two-variable solution which is as good or better than any other two-variable solution, and hence as good or better than any three-variable solution, or four-variable solution, or five-variable solution. We must, however, reiterate that the third basic solution may be, and in fact is, only one of many solutions which yield the maximum attainable income. This is readily seen from Table 5. There we note that a unit level of activity #2 can be substituted for its technically equivalent combination of activities #1 and #3 without either gain or loss in total income. (Again note the data of column 4.) Hence, we can substitute into the third basic solution any number of units of activity #2 for a like multiple of activity #2's technically equivalent combination of activities #1 and #3 (provided the level of either of the latter two activities is not forced below zero). Any such substitution yields a solution which provides the same maximum income. The reader may test this for himself.

The above sketches the basic procedures in the solution of a linear programming problem via the iterative process of the simplex method. Actually, a high-speed computing machine does not carry through the operations in the exact manner we have described. Essentially, however, the machine performs the same steps based upon the same logic. At times difficulties crop up in the solution, such as the problem of degeneracy. We do not look into these difficulties here since they involve rather technical discussion and since they have been fully discussed in the literature cited.

4. THE DUAL: A GRAPHIC ANALYSIS

Having presented the essence of linear programming by graphic analysis, and having illustrated the use of the simplex method in solving a linear programming problem, we have one remaining important aspect to discuss. This aspect concerns the dual of a linear program.

Mathematically speaking, every linear program expressed as a minimum has a dual problem expressed in terms of a maximum, and every linear program expressed as a maximum has a dual problem expressed in terms of a minimum. Generally in social science and regional analysis, the dual problem is artificial, but it does when solved yield meaningful results. Often its significance is not direct and immediately apparent, as will be seen in the following discussion. Nonetheless, it still can be extremely useful to examine the dual. New insights can be gained on the structure and interrelation of regional forces. A check on the formulation of the original linear program can be had. Finally, as will be demonstrated below, an easier and quicker way of arriving at the solution to the linear program may be possible.

To develop the notion of the dual, we may reconsider the problem of the preceding section. The region possesses the given quantities of the two resources, water and land, which are to be priced in such a manner as to lead to a wise program of utilization. Assume that among firms conditions of perfect competition prevail, that is, that there are many firms, and that each firm is free to engage in, or abandon, any productive activity and has unrestricted opportunity to bid for resources. (For any given activity, each firm faces the same set of input requirements as given by Table 2.) In such a situation it is clear that no activity will be operated by any firm at a profit (where profit per unit level is defined as the difference between income and costs per unit level).[11] For if a firm were operating an activity at a profit, new firms could begin to engage in this activity, would bid for the scarce water and land and thereby drive up the prices of these resources until all profits were eliminated. Hence, we may conclude that costs per unit level of an activity will be at least as great as income per unit level.

But what are costs? In this problem, they are simply the costs of water and land. (To be more realistic, the reader may suppose that there

[11] The term profit is used here in the sense of "excess" returns to management, over and above what is considered normal payment for capital, managerial ability and other factors which the owner of the firm may supply. See the following footnote for further clarification.

are other costs -- such as costs for raw materials, labor, capital services, and management ability -- which are constant and fixed per unit level of activity and which have already been deducted from gross sales value per unit level of activity to yield the income per unit level of activity.)[12] Since a_{11} units of water are required per unit level of activity #1, this amount multiplied by the price of water, namely, P_1 yield water costs $(a_{11}P_1)$ per unit level of activity #1. Also, since a_{21} units of land are required per unit level of activity #1, this amount multiplied by the price of land, namely, P_2 yields land costs $(a_{21}P_2)$ per unit level of activity #1. The sum of these two cost items, namely, $a_{11}P_1 + a_{21}P_2$ represents costs per unit level of activity #1, which by the assumption of perfect competition must not be less than c_1, the income per unit level of activity #1.

Similarly, we derive that $a_{12}P_1 + a_{22}P_2$ (the cost per unit level of activity #2) must not be less than c_2 (the income per unit level of activity #2), and that $a_{13}P_1 + a_{23}P_2$ (the cost per unit level of activity #3) must not be less than c_3 (the income per unit level of activity #3). Thus the constraints which our assumptions impose and within which we must keep are:

$$
\begin{aligned}
a_{11}P_1 + a_{21}P_2 &\geq c_1 \\
a_{12}P_1 + a_{22}P_2 &\geq c_2 \\
a_{13}P_1 + a_{23}P_2 &\geq c_3
\end{aligned}
\tag{10}
$$

For any activity (whether engaged in or not), unit level costs must equal or exceed unit level income.[13]

In Table 2 are given the values for the **a** coefficients. We put these values into inequalities (10) as well as the values for c_1, c_2, and c_3. We obtain:

$$
\begin{aligned}
0.5P_1 + 0.2P_2 &\geq \$1.00 \\
0.6P_1 + 0.15P_2 &\geq \$1.00 \\
0.7P_1 + 0.1P_2 &\geq \$1.00
\end{aligned}
\tag{11}
$$

We now wish to graph what these relations would be if only the equality held i.e., if the sum of terms on the left-hand side of each equation exactly

[12] As noted earlier, we must assume that prices are fixed not only for any commodity produced by an activity, but also for any of its inputs **excluding** scarce resources, but including such items as capital services, management services, raw materials, and labor (when labor is not considered a **scarce** resource in the problem). Since a set of coefficients is also posited which indicates constant outputs and inputs per unit level of any activity, the supposition suggested in the text is in keeping with the nature of a customary linear program.

[13] Thus, in the typical situation when the equality holds for one or more of these constraints (i.e. unit level costs = unit level income), the inequality will hold for one or more other constraints (i.e. unit level costs will exceed unit level income). The activities corresponding to these latter constraints will not be operated.

equalled $1.00. On Figure 2 lines AB, CD and EF represent these relations in their equality form, referring respectively to activities #1, #2 and #3. [The prices of water and land (P_1 and P_2) are measured along the horizontal and vertical axes, respectively. For the moment we ignore the dashed lines.] Thus we can say that all points which represent combinations of values for the prices of land and water which satisfy the first constraint (i.e. yields unit level cost as least as great as unit level income for activity #1) either lie on line AB or above and/or to the right. Likewise, only points on or above and/or to the right of line CD represent values for the prices which satisfy the second constraint (i.e., yields unit level cost equal or greater than unit level income for activity #2). Finally, only points on or above and/or to the right of line EF represent values for the prices of land and water which satisfy the third constraint. All together, therefore, only points lying on or above and/or to the right of the segmented line EMB represent combinations of prices which satisfy all three constraints; by the assumption of perfect competition they are the only combinations from which we are allowed to choose.[14]

Now consider the choice of a set of prices for the resources, water and land. Let us select any one from the set represented by points above and/or to the right of the segmented line EMB. For ease of exposition, let us select the price of $2.00 for water and $10.00 for land, as given by point K. If all the stock of water and land were employed, total returns to the owners of these resources, represented by W where

$$W = P_1R_1 + P_2R_2 \tag{12}$$

would be $30MM since from above R_1 = 6MM units of water and R_2 = 1.8MM units of land. Note that point K lies on a $30MM line, which is a locus of points representing all combinations of water and land prices which would yield $30MM provided 100 per cent of the stock of both water and land were utilized. But this total returns figure is purely artificial. And henceforth we shall designate W as total **fictitious** returns. For if the price of water were $10.00 and that of land $2.00 (or if these prices were represented by any other point on the $30MM line) unit level costs in each of the three activities would well exceed unit level income. Each activity would have to operate at a loss. There would be in fact no production, and hence no employment of resources. Total **effective** returns to resource owners would be zero, and all resources would be idle. This situation is, of course, neither to the interest of the owners of the resources (whom we assume to be many in keeping with our framework of perfect competition) nor consistent with the objective of any regional planning authority interested in the wise use of resources.

[14] Suppose we were to consider a combination of prices which is not represented by one of these points. Suppose this combination is given by point J in Figure 2. The combination satisfies the first two constraints, but not the third constraint. Consequently, if prices as given by J had been in effect, unit level costs of activity #3 would have been less than unit level income, i.e. there would have been a profit from engaging in activity #3. This would have led to expansion of the level of activity #3 by both new and old firms such that resource prices would have been bid up and the profit in activity #3 would have been reduced to zero. Thus point J represents a set of prices inconsistent with an equilibrium state under assumptions of perfect competition. We are not permitted to consider this combination.

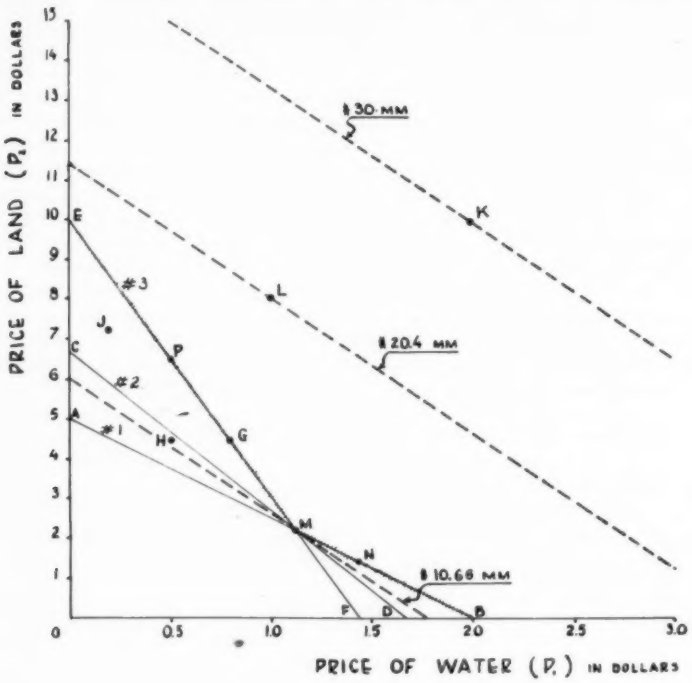


FIGURE 2.

We therefore seek a more desirable set of prices. Suppose we consider a set of lower prices, say $8.00 for land and $1.00 for water. (Such prices might be established by a regional planning authority, or might result from competitive undercutting by the many owners of the resources.) These prices are represented by point L in Figure 2. Immediately it is seen that it still would not be profitable for any firm to engage in any activity. No resources would be employed. Total **effective** returns to resource owners would be zero, although total **fictitious** returns (i.e. the income W of equation (12) which would be realized if 100 per cent of the stock of water and land were sold at the ruling prices) would be **$20.4MM**, having previously been $30.00 **MM**.

It is clear that the two prices must fall still more for production to occur. They must fall sufficiently so that they can be represented by some point on the segmented line **EMB**. (At any set of prices which can be represented by a point above and/or to the right of this segmented line, each activity operates at a loss.) Suppose we consider the set of prices, $6.50 for land and $0.50 for water, represented by point **P**, which set might have resulted from action by a regional planning authority or from the competitive process of price cutting. At this set of prices activity #3 can be carried on without loss; while activities #1 and #2 can be operated only at loss and hence will not be pursued. We may presume that activity #3

would be operated at the highest level permitted by the availability of resources. Since the stocks of water and land are 6MM and 1.8MM units respectively, and since 0.7 and 0.1 units of water and land are required per unit level of activity #3, activity #3 can be pursued at a level of 6MM/0.7 or approximately 8.6MM units. At this level all the stock of water and 0.86MM units of the available 1.8MM units of land are utilized. Income received by the owners of the water resources and of the land resources are $3MM and $5.6MM respectively, a total of $8.6MM. Total fictitious returns (namely, W) falls again, this time to the level of $14.7MM.

The above situation represents an improvement over the previous one from the standpoint of both the regional planning authority and owners of resources. Since not all the land is used, is there still a better situation? Certainly, we could expect those owners of land who were unsuccessful in selling their resource (and thus who receive no income) to explore another situation. Suppose, to decrease the amount of idle land, the regional planning authority sets a lower price on land, say $4.4, or suppose this price is reached as a result of competitive price-cutting by land owners. The corresponding price of water would be raised to $0.8, as given by point G in Figure 2.[15] Once again, only activity #3 would be pursued; and if it were pursued at its maximum level, namely, 8.6MM units all the stock of water and 0.86MM units of the available 1.8MM units of land would be utilized. Income received by the owners of the land resource would have fallen to $3.8MM (because of lower land price) while income received by water owners would have increased to $4.8MM (because of higher water price). Total **effective** returns to resource owners would, however, not have changed, being $8.6MM; whereas total **fictitious** returns (namely, W) would have fallen to the level of $12.7MM.

We conclude that as a consequence of the price changes represented by the shift from point P to point G on Figure 2, the pattern of resource use does not change (the same amount of land being idle), total **effective** returns to resource owners does not change (although land owners obtain a smaller share), but total **fictitious** returns does continue to fall. We could examine other combinations of prices involving still lower prices for land (reflecting the desire to encourage fuller resource use, or the competitive process). But all combinations represented by points lying between points G and M, excluding point M, would not lead to any greater use of land, or any increase in total **effective** returns to resource owners, although the closer we approach M, the smaller the share of total **effective** returns which goes to land owners, and the smaller total fictitious returns. In general it can be said that all combinations of prices represented by points between E and M, excluding M, lead to the identical pattern of resource use and yield the same total **effective** returns to resource owners. The closer the point is to M, the smaller (larger) the share of **effective** returns received by land owners (water owners), and the smaller total

[15] If the price of water were not raised and were kept at $0.5 (with the price of land set at $4.4), we would be selecting a set of prices, as represented by point H, which would give rise to profits in both activities #2 and #3. But this is inconsistent with our assumption of perfect competition (profitless production). We are not permitted to select this set of prices.

In contrast, if the price of water were raised still higher, i.e. to a level greater than $0.8, then all activities could be carried on only at a loss. There would be no production, and no use of resources, a definitely inferior situation.

fictitious returns.

The situation abruptly changes for a set of prices represented by point M, where lines EF, CD and AB intersect. At such prices (namely $2.22 for land, and $1.11 for water), it is possible to pursue more than one activity without loss; in fact it happens that in this particular situation it is possible to pursue all three activities without loss. This provides much more flexibility in the use of resources (since each activity uses resources in different combinations -- see Table 2), and does lead to full use of land and water resources. Actually, each of an infinite number of combinations of levels of these three activities will lead to full use of resources. One such combination is 7.33MM units of activity #1 and 3.33MM units of activity #3; a second is 4MM units of activity #1 and 6.66MM units of activity #2; a third is 5.66MM units of activity #1, 3.33MM units of activity #2, and 1.66MM units of activity #3; etc.[16] Since each of these combinations fully utilize the resources, each yields to land owners effective returns of $4MM, and to water owners, $6.66MM. Total effective returns is thus $10.66MM, a significant increase over the effective returns yielded by any other point on the line segment EM (or by any other point above and/or to the right of it). In contrast, total fictitious returns falls to the lowest level yet reached, namely to $10.66MM, a level equal to total effective returns.

Is M the best point from the standpoint of both a regional planning authority and resource owners? Clearly, none of the resource owners will have any incentive to shave prices since he can sell all his stock at the ruling prices. (And if he should raise his price, no firm will purchase from him, since any firm doing so would operate at a loss.) But a regional planning authority might not be convinced. Suppose then we consider another set of prices, say $1.4 for water and $1.5 for land, as given by point N on stretch MB. (Recall that stretch MB pertains to activity #1, and not to activity #3 as stretch EM does.)[17] At these prices, activities #2 and #3 are not pursued, since to do so would involve loss. Activity #1 can be pursued without loss, and may be presumed to be operated at the highest level permitted by the stock of resources. Since 0.5 units of water and 0.2 units of land are required per unit level of activity #1 (see Table 2), activity #1 can be operated at a level of 1.8MM/0.2 = 9MM units, at which level the 1.8MM units of available land are fully used and only 4.5MM units of the available 6MM units of water are employed. Returns to land owners are $2.7MM and to water owners, $6.3MM, which yields total effective returns of $9.0MM. Total fictitious returns (namely, W) is $11.1MM. We therefore

[16]The infinite number of combinations of activities #1, #2 and #3 which will fully utilize the available resources can be obtained: (1) by starting with the combination of 7.33MM units of activity #1 and 3.33MM units of activity #3; and (2) by adding units of activity #2 (up to a maximum of 6.66MM) and simultaneously subtracting units of activities #1 and #3 at a rate of one-half unit of both #1 and #3 for every unit of #2 added. This is consistent with the data of Table 5.

[17]Also, to select a set of prices represented by a point below stretch MB would imply profits in activity #1 and be inconsistent with the assumption of perfect competition among firms; while to select a set of prices represented by a point above stretch MB would involve loss in the operation of every activity.

conclude that a shift from a set of prices given by **M** to a set given by **N** increases only total **fictitious** returns (the returns resource owners would receive provided they could sell all their stocks). Effective returns to both land owners and water owners fall, and thus total **effective** returns. Some of the stock of water is unutilized. Thus from the standpoint of both a regional planning authority and resource owners, the set of prices defined by **N** is less desirable than that defined by **M**. And if there are many small competitive water resource owners, we would expect price-cutting by those who were unsuccessful in selling their units, so that in effect an **M** set of prices would be realized if initially a **N** set were established.

In similar manner it can be shown that a set of prices represented by any other point along stretch **MB** is inferior to that represented by point **M** both in terms of achieving full use of resources and of maximizing total **effective** returns. Hence, **M** represents the desired set of prices. Here, full use of resources is achieved, total **effective** returns is maximized. Also, total **fictitious** returns is minimized at **M**. This is seen by studying the dashed lines of Figure 2. As already noted, each dashed line goes through all the points which represent combinations of prices which yield the total **fictitious** returns associated with that line. As we approach the origin the total **fictitious** returns associated with these parallel dashed lines declines. It would become zero at the origin where prices are zero; but we are not permitted to select such a set of prices. We are allowed to select only those combinations represented by points on, or above and/or to the right of segmented line **EMB**. Of all such points, **M** clearly lies on the lowest dashed line. Thus total **fictitious** returns is minimized at **M**, given our constraints; which is the same point at which total **effective** returns to resource owners is maximized and full use of resources is achieved. Thus, in our problem, we have shown that to find the set of prices which will maximize total **effective** returns to resource owners and achieve full use of resources is **to find that set of prices which minimizes an artificial concept, namely, total fictitious returns to resource owners** (i.e. **W** where by equation (12) $W = P_1R_1 + P_2R_2$) subject to the condition that for non-negative prices the costs of activities #1, #2, and #3 (i.e. $a_{11}P_1 + a_{21}P_2$; $a_{12}P_1 + a_{22}P_2$; and $a_{13}P_1 + a_{23}P_2$) not be less than -- or, put otherwise, be equal to or greater than -- unit level income (c_1, c_2 and c_3) respectively. This is the dual of the original linear program.

5. THE LINEAR PROGRAM AND ITS DUAL: INTERRELATIONSHIPS

We now set down side-by-side the original linear program and its dual. [Recall the linear program was to select a set of non-negative levels at which to operate each activity (i.e. X_1, X_2, and X_3) in order to maximize income (Y)[18] where the income from running each activity at unit level (i.e. c_1, c_2, and c_3 respectively) is \$1.00; subject to the constraints that the requirements of all three activities for land and water do not exceed the available stocks (R_1 and R_2).]

[18]Also recall that $Y = \$1 \times Z$ where $Z = X_1 + X_2 + X_3$; and that in equation (5) Y may also be viewed as the sum of the first three terms since by definition $c_4 = c_5 = 0$.

Linear Program

To maximize: $Y = c_1X_1 + c_2X_2 + c_3X_3$

subject to:

$$a_{11}X_1 + a_{12}X_2 + a_{13}X_3 \leq R_1$$

$$a_{21}X_1 + a_{22}X_2 + a_{23}X_3 \leq R_2$$

$$X_1 \geq 0;\ X_2 \geq 0;\ X_3 \geq 0$$

Dual

To minimize: $W = P_1R_1 + P_2R_2$

subject to:

$$a_{11}P_1 + a_{21}P_2 \geq c_1$$

$$a_{12}P_1 + a_{22}P_2 \geq c_2$$

$$a_{13}P_1 + a_{23}P_2 \geq c_3$$

$$P_1 \geq 0;\ P_2 \geq 0$$

First, we note that the solutions to the two problems yield **Y = W = $10.66MM**, i.e. the maximum income achievable turns out to be equal to the minimum realizable total **fictitious** returns to resource owners (which is equal to maximum total **effective** returns to resource owners).[19]

Second, the constants, c_1, c_2, and c_3 which respectively multiply the "choice" variables in the objective function of the linear program are each a limiting constant in a constraint of the dual. Therefore, there are as many constraints in the dual as there are choice variables in the linear program. At the same time, the limiting constants of the constraints of the linear program, namely, R_1 and R_2, are the coefficients which multiply the choice variables of the dual. Hence, there are as many choice variables in the dual as there are constraints in the linear program.

Third, the coefficients of the set of constraints of the linear program, **when transposed**, are the set of coefficients of the constraints of the dual. Or, put otherwise, a row of coefficients in a single constraint of a linear program, (such as a_{11}, a_{12} and a_{13}) becomes a corresponding column which multiplies the single corresponding choice variable (such as P_1) in the dual, and vice-versa; and any column of coefficients in the linear program (such as a_{12} and a_{22}) which multiply a single choice variable (here X_2) becomes the corresponding row in the set of constraints of the dual, and vice-versa.

Fourth, the sense of the inequalities in the constraints of the linear program and dual are the reverse of each other. For example where one is "less than" (<) as is the case for our linear program (and as must be the case for every maximum problem), the other is "greater than" (>) as is the case for our dual (and as must be the case for every minimum problem).[20]

[19] Thus, the linear program and its dual have a common solution here. It can be generally stated that a linear program and its dual have either a common solution or no solution at all.

[20] Thus, if the linear program is a minimum problem, the sense of the inequalities is "greater than", while the sense of the inequalities in the dual (which must be a maximum problem) is "less than".

This rule holds except that the inequalities pertaining to non-negative values for the choice variables in both the linear program (the X's) and dual (the P's) have the same sense.

Having presented the dual, its solution, and its relation to the linear program,[21] we can reexamine the usefulness of the dual. Clearly the magnitude to be minimized in our dual tends to be an artificial one, as is generally true of the magnitude to be minimized or maximized in the dual of any meaningful linear program.[22] Yet, posing and solving the dual permit deeper insight into the operation of the system and into the structure of the problem being examined. Useful, too, is the check on the formulation of the linear program which the formulation of the dual may afford, since certain corresponding relations must exist between a linear program and its dual. Perhaps, the most important use of the dual, however, is to ease the computation task. In the previous section, the simplex computation was developed for the linear program examined. Suppose we were to introduce a fourth possible activity into the linear program -- an activity which yields an income of $1.10 per unit level of activity and which requires 0.40 units of water and 0.24 units of land per unit level of activity.

The linear program becomes:

To maximize:

$$Y = c_1X_1 + c_2X_2 + c_3X_3 + c_4X_4 \tag{13}$$

subject to:

$$\begin{aligned} a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + a_{14}X_4 &\leq R_1 \\ a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + a_{24}X_4 &\leq R_2 \\ X_1 \geq 0 \quad X_2 \geq 0 \quad X_3 \geq 0 \quad X_4 &\geq 0 \end{aligned} \tag{14}$$

It is not possible to arrive quickly at the optimal solution to this new linear program via the simplex method. A set of computations would have to be made, as were performed in connection with Tables 3, 4 and 5. In contrast to this relatively time-consuming task, we can convert the new linear program into its dual, which involves simply the addition to constraints (11) of one more constraint, namely, that for activity #4 unit level cost be at least as great as unit level income (i.e. $0.40P_1 + 0.24P_2 \geq \$1.10$). If we add to our graph this new constraint in its equality form, as designated by line UV in Figure 3,[23] we directly observe that total fictitious returns to resource owners is minimized at the combination of resource prices given by point Q. This immediately indicates that the optimal set of prices of water and land are, in round numbers, $1.02 and

[21] It is interesting to note here that the "dual" of the dual is the original linear program, as the reader may verify for the situation we have examined.

[22] E.g., see Allen, op. cit., pp. 539-541; and Dorfman, Samuelson and Solow, op. cit., chaps. 3, 5, and 7.

[23] Figure 3 reproduces lines AB, CD and EF of Figure 2, which indicate the first three constraints of the dual.

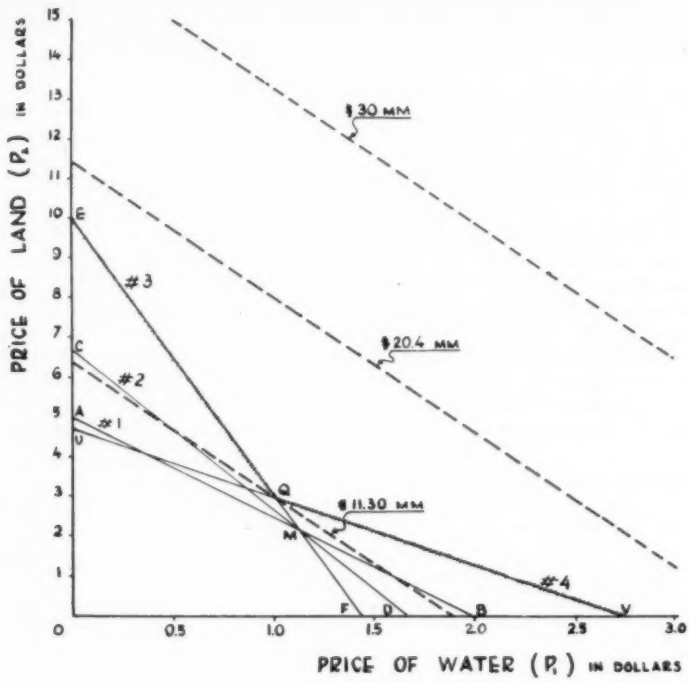


FIGURE 3.

$2.89, respectively. Knowing prices, we calculate total **fictitious** returns to resource owners to be approximately **$11.30MM.** This figure also represents the maximum total **effective** returns to resource owners, and is also equal to the maximum of income achievable in the original linear program, i.e. is equal to the total income which would be yielded by the optimal solution of the linear program. Also we note from Figure 3 that at the optimal prices (as given by point Q), the operation of both activities #1 and #2 would incur loss. Therefore, neither will be operated. In contrast, activities #3 and #4 can be operated without loss. We therefore let $X_1 = X_2 = 0$ in the two constraints (14), use the equality form of these constraints and solve the resulting two equations in two unknowns for X_3 and X_4. We find that $X_3 = 5.63$MM and $X_4 = 5.16$MM; and that accordingly by equation (13), Y = $11.30MM.

In the above manner, we quickly arrive at the solution of the new linear program by first solving its dual. This indeed makes the dual a useful device. But this case is a special case and unlike the general situation. In the general situation where a linear program has more than two constraints, the graph of the dual will have more than two dimensions; and hence a quick graphic solution to the dual is not possible. Generally speaking, one chooses to solve the linear program or its dual, depending

on which has the fewer constraints.[24]

The preceding problem, embracing four possible activities, also points up certain other relations between a linear program and its dual. When any choice variable of a linear program is at a positive level in the optimal solution, the corresponding constraint in the dual solution is binding (that is satisfied with exact equality). Thus we find that both X_3 and X_4 of the above new linear program are positive in the optimal solution; concomitantly, constraints 3 and 4 are met with exact equality in the dual solution since point Q in Figure 3 lies on both lines EF and UV (i.e. unit level cost equal unit level income for both activities #3 and #4). In parallel fashion, when any choice variable of the dual is at a positive level in the optimal solution, the corresponding constraint in the linear program solution is binding (that is satisfied with exact equality). Thus we find that both P_1 and P_2 of the dual are positive; and correspondingly the two constraints of the linear program are met with exact equality, i.e. the supplies of the two resources, water and land, are both fully utilized.

Furthermore, if a choice variable in the linear program takes the value zero in the optimal solution, then the corresponding constraint in the dual solution will usually be satisfied with an inequality.[25] Thus, X_1 and X_2, the levels of activities #1 and #2, are zero in the optimal solution of the above new linear program; correspondingly, the constraints of the dual relating to activities #1 and #2 are not binding (i.e. satisfied with an inequality) in the solution to the dual. This latter relation is portrayed on Figure 3 where point Q, representing the optimal set of prices, lies above both lines AB and CD, which signifies that for both activities #1 and #2 unit level cost exceeds unit level income. In parallel fashion, if a choice variable in the optimal solution to the dual is zero, then the corresponding constraint in an optimal solution to the linear program will

[24]This general rule, however, does not hold where a graphic solution is possible as above; in the above problem, the dual has the greater number of constraints.

[25]The qualification "usually" must be made since in situations where more than one optimal solution to the linear program exists --- as is the case for the "three-activity" linear program which was first examined --- a choice variable may be set at zero in one optimal solution, and yet the corresponding constraint in the dual will be satisfied with an equality. For example, in the linear program of three activities, to which Tables 3, 4, and 5 refer, and to whose dual Figure 2 pertains, one optimal solution is $X_1 = 7.33MM$; $X_3 = 3.33MM$; and $X_2 = 0$. Since $X_2 = 0$, we would normally expect the corresponding constraint in the dual to be satisfied with an inequality. Actually, it is satisfied with an equality as indicated in Figure 2 by the fact that line CD, which refers to activity #2, passes through point M, the point representing the optimal set of prices for the dual problem, and which therefore indicates for activity #2 that unit level costs equals unit level income. When such a situation obtains, it signifies that the activity which is run at zero level in the given optimal solution could be operated at a positive level in one or more other optimal solutions. (Recall that an infinite number of optimal solutions may be possible in the sense that all such solutions yield the identical, maximum income.) This is in line with the conclusion already reached that the above combination of X_1 and X_3 is only one of an infinite number of optimal solutions for the three activity problem considered. In each of these other solutions, activity #2 would be run at a positive level.

usually be satisfied with an inequality.[26] For example, if we were to identify the optimal solution to the dual which corresponds to the linear program represented on Figure 1 above, we would find that the price of both labor and land is zero. Accordingly, the constraints on the use of labor and land can normally be expected to be satisfied with an inequality. This expectation is confirmed by the fact that on Figure 1, the optimal solution represented by point T lies below both RP, the labor limitation line, and LW, the land limitation line. This relationship indicates that the supply of labor and land are more than sufficient; therefore, some units of both land and labor must be idle which situation is consistent with the zero prices for land and labor.

These rules, which are extremely useful in proceeding from an optimal solution for a linear program to the optimal solution for its dual, and vice versa, can also be stated in converse fashion. When for an optimal linear program solution, a constraint is found binding (met with an equality), the corresponding choice variable in the optimal dual solution is usually positive.[27] When for the optimal linear program solution, a constraint is found non-binding, the choice variable in the optimal dual solution is always zero. Likewise, when a constraint in an optimal dual solution is binding, the choice variable in the optimal linear program solution will usually be positive. When the constraint is non-binding, the choice variable in the optimal linear program solution will be zero.

6. INTERMEDIATE COMMODITIES

Before we can generalize the scheme thus far developed, we must consider intermediate commodities. To do sp adequately, it is convenient to formulate in a more comprehensive manner the concept of an activity, and to alter our notation.

Hitherto, we have implicitly assumed that the **output** from the operation of an activity at unit level was marketed at a fixed price and yielded to the region $1.00 income (after allowance for other fixed unit costs). We must now make explicit this output. Let there be associated with each activity a set of both inputs and outputs. If we number commodities (including resources, services, etc), from 1 to s, then we can represent the inputs and outputs of an activity by a set of coefficients. We may let the set a_{11}, a_{21}, a_{31},...,a_{s1} denote the amounts of commodities 1, 2, 3,..., and finally s associated with the operation of activity #1 at unit level. For a typical activity, most of these coefficients will be zero. For example, if activity #1 is wheat growing, and the fth commodity is iron ore, we would expect the coefficient a_{f1} to be zero, indicating that the input (or output) of iron ore per unit of the wheat growing activity is zero. When there are outputs associated with an activity, the relevant coefficients will be positive. For example, if wheat and hay are products of the wheat growing activity and if wheat is commodity #1 and hay commodity #8, then both the coeffic-

[26] However, as observed in the previous footnote, the constraint will be satisfied with an equality, if there are more than one optimal solution to the dual, and for which solutions the choice variable considered may be non-zero.

[27] Thus, in the four activity linear program discussed immediately above, the constraints on the use of both water and land are binding; and as to be expected, the prices of both water and land are positive as indicated by point Q in Figure 3.

ients a_{11} and a_{81} will be positive. When there are inputs associated with an activity, the relevant coefficients will be negative. For example, water, land, labor, and capital may all be required as inputs by the wheat growing activity. If they are classified respectively as commodities, g,h, k and m, then the coefficients a_{g1}, a_{h1}, a_{k1} and a_{m1} are negative. **Note that by this convention, we change the sign of the coefficient associated with an input. In the previous sections, positive coefficients were used to denote inputs.** Such a convention was desirable in treating the simple cases which were involved. In the development of a more general statement on linear programming, it becomes highly desirable to change the previous convention and **treat inputs as negative items.**

If we now consider many possible activities, which number from 1 to n, we may say that associated with any activity say activity j, there are a set of coefficients, a_{1j}, a_{2j},...,a_{sj} which indicate inputs and outputs per unit level of activity j. For each activity, it becomes convenient to list these inputs and outputs in a column as follows:

$$\begin{bmatrix} a_{11}, a_{12}, \ldots, a_{1j}, \ldots, a_{1n} \\ a_{21}, a_{22}, \ldots, a_{2j}, \ldots, a_{2n} \\ a_{31}, a_{32}, \ldots, a_{3j}, \ldots, a_{3n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ a_{s1}, a_{s2}, \ldots, a_{sj}, \ldots, a_{sn} \end{bmatrix}$$

The first column refers to the set of inputs and outputs associated with the first activity, the second with the second activity, the jth with the jth activity, and finally the nth column with the nth activity. Such a listing of coefficients by columns according to the order of both the commodity and activity classification may be termed a matrix, an activity coefficient matrix. As already suggested, most of these coefficients will be zero. Also as one moves across any one row, say the ith which refers to commodity i, he will note that for some activities the coefficient is negative, indicating that these activities use i as an input while for others the coefficient is positive, indicating that these latter produce i. If i is a good such as iron ore, most of the coefficients along this row would be zero. If i is a good such as power, most of the coefficients would be negative, though a few would be positive, and some zero.

If we now restate the above linear programming problem which involved three activities and two resources -- water, which we will designate commodity g; and land which we designate commodity h -- the constraints appear as:

$$-a_{g1}X_1 - a_{g2}X_2 - a_{g3}X_3 \leq R_g$$

$$-a_{h1}X_1 - a_{h2}X_2 - a_{h3}X_3 \leq R_h$$

Note that all the above coefficients ($a_{g1},\ldots,a_{h3}$) are negative since they represent inputs; we thus multiply each by -1 to convert them into the positive numbers, $-a_{g1},\ldots,-a_{h3}$. Multiplying X_1, X_2, and X_3 by the positive numbers, $-a_{g1},\ldots,-a_{h3}$ permits us to state the constraints properly, i.e., to state that the sum of the amounts (positive) of water and land required by each activity be equal to or less than the available supply of water and land, respectively.

Into this framework we can now introduce an intermediate commodity, say i. This commodity may be produced singly or jointly by one or several activities.[28] It may be a commodity, such as steel ingots, which is used solely in industrial production, or it may be a commodity, such as transportation, which is demanded as an intermediate by producers but also as a finished product by households. To simplify our exposition we examine a case where the intermediate commodity: (a) is produced by a single activity only, and is the only output of this activity; and (b) is consumed by only the three activities of our problem.[29] We designate the activity which produces this intermediate commodity i as activity #4.

The objective function of the linear program is not basically altered. For activity #4, income per unit level is zero, since it can always be shown that the price of the intermediate i must exactly equal the cost of the required inputs to produce a unit of i.[30] Thus the term c_4X_4 is zero, and income realized still equals $c_1X_1 + c_2X_2 + c_3X_3$. However the set of constraints changes. They become:

$$-a_{g1}X_1 - a_{g2}X_2 - a_{g3}X_3 - a_{g4}X_4 \leq R_g$$

$$-a_{h1}X_1 - a_{h2}X_2 - a_{h3}X_3 - a_{h4}X_4 \leq R_h$$

$$-a_{i1}X_1 - a_{i2}X_2 - a_{i3}X_3 - a_{i4}X_4 \leq 0.$$

To the first constraint has been added the term, $-a_{g4}X_4$, which represents the water requirements of activity #4, the coefficient $-a_{g4}$ representing the amount of water required per unit level of activity #4 and X_4 representing the level at which this intermediate commodity producing activity is operated. Thus the first constraint states that the requirements of water by the first three activities plus the requirement of water by the new activity #4 must not exceed the available supply.

To the second constraint has been added the term, $-a_{h4}X_4$, which represents the land requirements of activity #4. The constraint now reads: the

[28] When it is produced jointly by the assumption of constant production coefficients, it must be produced in fixed proportion with other outputs.

[29] As will be discussed below, a commodity which is consumed both by industry (as an intermediate) and by households (as a finished product) can best be treated as two commodities. In the list of commodities from 1 to s, it appears once as an intermediate, and a second time as a finished product.

[30] See below. The reader is also referred to the general literature cited for discussion of this point.

requirement of land by all four activities must not exceed the available supply of land.

Finally, we have added a third constraint which essentially states: the sum of the requirements of the intermediate commodity I must not exceed the output of I; or, conversely, the output of I must be at least as great as the sum of the requirements of I by all activities. The first three terms of the constraint represent the consumption of I by activities #1, #2, and #3, respectively; since these are the only activities consuming I, the sum of the first three terms yields total requirements of I. The fourth term, $-a_{i4}X_4$, is the **negative** of the total output, since a_{i4} is a positive coefficient indicating the output of I per unit level of activity #4 and X_4 is the level at which activity #4 is operated. Thus the constraint reads that when total output $a_{i4}X_4$ is subtracted from total requirements $(-a_{i1}X_1 - a_{i2}X_2 - a_{i3}X_3)$,[31] we obtain zero (in which case output would have equalled total requirements) or a negative number (in which case output would have exceeded total requirements).[32] This is essentially what the first sentence of this paragraph states.

7. A GENERALIZED SINGLE REGION MODEL

We now have presented the linear program in simple graphic and algebraic forms, developed the simplex method for a three activity problem, formulated and solved the dual of this problem, sketched the more important relationships between the linear program and its dual, and finally introduced intermediate commodities. All this was done for programs which involved from two to four activities, and from two to four resources in limited supply. At this point we may fruitfully generalize. (However, the materials in this section are more difficult. The reader may prefer to skip this section,[33] and go on to the next section. The next section treats a simple interregional model and is easier to digest.)

Let there be for our hypothetical isolated region many, say **n** possible activities for achieving new income. The region confronts many, say **m** possible constraints which reflect fixed supplies of resources available for generating new income, limitations on the use of intermediate commo-

[31] Recall that a_{i1}, a_{i2}, a_{i3} are negative coefficients representing inputs; hence $-a_{i1}$, $-a_{i2}$, and $-a_{i3}$ are positive numbers.

[32] In our simplified problem the third constraint could be written in the equality form only, i.e.,

$$-a_{i1}X_1 - a_{i2}X_2 - a_{i3}X_3 - a_{i4}X_4 = 0.$$

We know that only activity #4 produces I, and that this activity yields zero income per unit level while at the same time employs resources, some of which are scarce. Therefore excess production of I would be inconsistent with an optimal solution.

More generally, however, there could be an excess production of I if I were a joint product of one or several activities. In such a situation the inequality could hold (or, a zero-cost disposal could be introduced to dispose of the surplus I).

[33] This section might well be read after the following two sections.

dities, and other restrictions on commodity consumption.[34] Order the commodities associated with these constraints from 1 to m. Assume as given the prices of finished products, i.e. of commodities (m+1) to s, where as above commodities run from 1 to s. (Clearly this assumption is most valid when the quantities of inputs and outputs to be associated with a linear program are very small relative to the inputs and outputs of the existing industrial structure. As these quantities increase in relative size, the reality of the assumption for a regional economy decreases.) The linear program and its dual become:

Linear Program	Dual
Max: $Y = c_1X_1 + \ldots + c_nX_n$	Min: $W = P_1R_1 + P_2R_2 + \ldots + P_mR_m$
subject to:	subject to:
$-a_{11}X_1 - a_{12}X_2 - \ldots - a_{1n}X_n \leq R_1$	$-a_{11}P_1 - a_{21}P_2 - \ldots - a_{m1}P_m \geq c_1$
$-a_{21}X_1 - a_{22}X_2 - \ldots - a_{2n}X_n \leq R_2$	$-a_{12}P_1 - a_{22}P_2 - \ldots - a_{m2}P_m \geq c_2$
.....	.....
$-a_{m1}X_1 - a_{m2}X_2 - \ldots - a_{mn}X_n \leq R_m$	$-a_{1n}P_1 - a_{2n}P_2 - \ldots - a_{mn}P_m \geq c_n$
$X_1 \geq 0; X_2 \geq 0, \ldots, X_n \geq 0$	$P_1 \geq 0; P_2 \geq 0, \ldots, P_m \geq 0$

In this linear program and its dual each term on the left hand side of each constraint has been multiplied by -1. As indicated in the previous section, this step is necessary in order to retain the meaning of each constraint since except for coefficients representing intermediate outputs each of the above coefficients represents an input and consequently by our new definition is negative. Thus, for example, $-a_{12}X_2$ and $-a_{2n}P_2$, are each positive magnitudes, one representing a positive consumption, the other a positive cost. In the case of coefficients indicating outputs of an intermediate, multiplying by -1 is appropriate since in the relevant constraint the output of an intermediate is to be subtracted from the total requirement for that intermediate.

In the above statement, R_i $(i = 1,\ldots,m)$[35] is the fixed available stock of resource i, or other limiting magnitude. In the case of an intermediate commodity, this limiting magnitude is zero. Also c_j $(j = 1,\ldots,n)$[36] represents the per unit level income of activity j. In the case where ac-

[34] These latter restrictions will be discussed in a later section.

[35] R_i $(i = 1,\ldots,m)$ represents the series of constants, $R_1, R_2, \ldots, R_m$.

[36] c_j $(j = 1,\ldots,n)$ represents the series of constants, $c_1, c_2, \ldots, c_n$.

tivity j uses as inputs one or more commodities numbered from 1 to m, and produces only one commodity, say m+1,

$$c_j = a_{m+1,j} \cdot \bar{P}_{m+1}$$

i.e. is equal to the per unit level output of commodity m+1 times its price which is already specified.[37] Where activity j uses as inputs commodities numbered from 1 to m, and produces only two finished products, say goods m+1 and m+2,

$$c_j = a_{m+1,j} \cdot \bar{P}_{m+1} + a_{m+2,j} \cdot \bar{P}_{m+2}$$

i.e. is equal to the sum of the per unit level output of m+1 times the price of m+1 plus the per unit level output of m+2 times the price of m+2.

Generally speaking, where many final products are associated with activity j, we have

$$(15) \qquad c_j = a_{m+1,j} \cdot \bar{P}_{m+1} + a_{m+2,j} \cdot \bar{P}_{m+2} + \ldots + a_{sj} \cdot \bar{P}_s = \sum_{q=m+1}^{s} a_{qj} \cdot \bar{P}_q$$

i.e. the per unit level income of the jth activity, given the prices for commodities (m+1) to s, is equal to the sum of the products resulting from multiplying each per unit level output (positive) of a finished product by its price.[38] Per unit level of the jth activity this yields total sales values, i.e. income. Typically, all but a few of the a_{qj} will be zero signifying that a single activity produces only a relatively few, if more than one, of the many possible final products.[39]

In the case of activities which produce only intermediate commodities (which are commodities included in the group of commodities numbered from $1,\ldots,m$) income per unit level, namely c_j, is zero since all the coefficients, $a_{m+1,j},\ldots,a_{sj}$ of eq. (15) are zero. For such an activity the corresponding constraint in the dual is:

$$(16) \qquad -a_{1j}P_1 - a_{2j}P_2 - \ldots - a_{jj}P_j - \ldots - a_{mj}P_m \geq c_j$$

[37] Since this price is given beforehand, we place a bar over its symbol to indicate this, and to distinguish it from other prices which are not given beforehand and which are to be determined from the operation of the model. These latter are price variables and are unbarred.

[38] Recall that if an activity uses as input an item such as milk, which in reality is consumed by both industry (baking) and households, it is **by convention** using as input the commodity, **milk (as intermediate).** This commodity is listed among the first m commodities, while the commodity, **milk (as finished product)**, is listed among the commodities which run from m+1 to s. As a consequence, none of the coefficients $a_{m+1,j},\ldots,a_{sj}$ can be negative.

[39] In addition, it may produce intermediate commodities, which production would be listed in the first m of the a coefficients associated with this activity. (As will be observed below, an activity can produce an item only as an intermediate if that item is consumed by both industry and households.)

Take the term $a_{jj}P_j$ over to the right. Also assume that the unit level of activity j has been defined so that $a_{jj} = 1$ and that activity j produces only one commodity. We have, since $c_j = 0$:

$$(17) \quad -a_{1j}P_1 \quad ,\ldots, \; -a_{j-1,j}P_{j-1} - a_{j+1,j}P_{j+1} \quad ,\ldots, \; -a_{mj}P_m \geq P_j$$

That is, the per unit level cost of the first input $(-a_{1j}P_1)$ + the per unit level cost of the second input $(-a_{2j}P_2)$ +,...,+ finally the per unit level cost of the mth input $(-a_{mj}P_m)$ is greater than or equal to the price of the intermediate commodity j. We know that when activity j is actually operated, the equality holds for (17); and hence when the intermediate commodity j is produced its price must be exactly equal to the sum of all per unit level input costs. Thus an optimal solution to a linear program and its dual yields not only prices (rents) for all resources (services) in limited supply but also prices for all intermediate commodities.[40]

In general, the discussion of the preceding sections concerning the interrelationships of a linear program and its dual, and the derivation of an optimal solution may be taken to apply to the above generalized regional model.

8. A SIMPLE INTERREGIONAL MODEL

At this point we are in a position to start the construction of interregional models. Our procedure will again be to develop a model for a very simple case, present and discuss the basic structure of interregional models by reference to this elementary model and certain extensions of it, and finally generalize.

Let there be a closed interregional system, consisting of two regions only, region A and region B. Each region may produce only three commodities: textiles (consumed by households only), coal (consumed by both industry and households) and transportation (consumed by industries only, each region being responsible for the delivery of its product to other regions). Available for production are resources, the only one in limited supply in any region being labor. We posit that R_1^A and R_1^B are the fixed supplies of labor available in regions A and B, respectively. Let the objective be to maximize the sum of the regional incomes.

It is convenient to set down systematically the list of commodities and set of activities. We bear in mind that any item in one region is a different commodity than the same item in a second region. Textiles in region A is a different commodity in our system than textiles in region B. This is

[40] Where activity j produces two or more strictly intermediate commodities -- a case of joint production which we generally do not consider in this elementary statement -- then the two or more corresponding terms of relation (16) must be brought over to the right-hand side. If the resulting supply of all but one of these intermediates is in excess (exceeds requirements), the price of all but the one not in excess supply will be zero, and the price of this latter intermediate can be derived from the equality form of (17). If the resulting supply of two or more of these intermediates is not in excess, additional relation of the type (17) are required if a unique set of prices for the intermediates involved is to be derived.

reflected by the fact that the price of textiles in region A will usually be different from the price of textiles in region B; and this difference need not be equal to the cost of transporting one unit of textiles from one region to the other. Likewise, coal, transportation and labor in region A are different commodities from their respective counterparts in region B. Consequently there are eight different commodities in our simple case. However, two of these commodities, namely coal in region A and coal in region B, are used as both intermediates (by industry) and finished product (by households). In our case -- and typically in interregional linear programming -- each such commodity must be viewed as two commodities, i.e. we must replace the commodity coal in region A by two commodities, namely, coal (as intermediate) in region A and coal (as finished product) in region A. Hence our model comprises ten commodities, as listed in the left-hand tab of Table 6.[41]

Just as each commodity has a regional specification, so does each activity in our model. Thus, as recorded in Table 6, textile production in region A (the first activity in this region) is a different activity than textile production in region B (the first activity in region B); and likewise with transportation production and coal production, which are respectively the second and third activity in each region. In each of these production activities, we encounter the usual input-output structure. Going down the first column of Table 6, which lists the relevant coefficients **per unit level** of activity #1 in region A, we note an input of labor in region A, as designated by a^{A}_{11}. The first subscript of this coefficient indicates the number of the commodity being used as input; the second subscript indicates the number of the activity using this input; and the superscript indicates the regional specification of both the commodity and the activity. [We shall examine below the case where the input (or output) and activity have different regional specifications.] The second coefficient in the first column i.e., a^{A}_{21} indicates the input of the commodity #2,A (i.e. transportation in region A) into the activity #1,A (i.e. textile production in region A). The third coefficient a^{A}_{31} represents the input of coal in region A per unit level of activity of #1,A. The fourth coefficient is 0, indicating that coal (as finished product) in region A is neither an input or output of textile production in region A. The fifth coefficient in column 1 is a^{A}_{51}. This coefficient indicates the output of textiles in region A per unit level of textile production in region A. Usually, it is convenient to define a unit level of textile production in region A so that a^{A}_{51} is unity. Following our convention this coefficient, representing an output, is positive; while the other non-zero coefficients in column 1, each representing an input, are negative. Observe, too, that coefficients assoc-

[41] Note that we have exactly the same items in each region. This convention is not at all necessary. We can have different types of resources in each region, different items produced in each region, as well as different numbers of commodities in each region. We nonetheless maintain regional symmetry in the model since doing so facilitates the exposition.

Note, too, that we have constructed the model so that it embraces each type of commodity generally encountered in interregional linear programming. We have: (a) a resource (labor) in each region which may or may not be mobile; (b) a pure intermediate (transportation) in each region; (c) an item (coal) which behaves both as an intermediate and a finished product; and (d) a commodity (textiles) which is strictly a finished product (i.e. for consumption by households only).

TABLE 6. **Activity Coefficient Matrix: Simple Interregional Model.**

COMMODITIES \ ACTIVITIES	#1,A Textile Production in A	#2,A Transportation Production in A	#3,A Coal (Intermediate) Production in A	#4,A Coal in A: Shift to Finished Product Status (Dummy)	#5,A Shipment of Transportation from A to B	#6,A Shipment of Coal (as Intermediate) from A to B	#7,A Shipment of Textiles from A to B
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
#1,A Labor in A	a^{A}_{11}	a^{A}_{12}	a^{A}_{13}	0	0	0	0
#2,A Transportation in A	a^{A}_{21}	a^{A}_{22}	a^{A}_{23}	0	a^{A}_{25}	a^{A}_{26}	a^{A}_{27}
#3,A Coal (as Intermediate) in A	a^{A}_{31}	a^{A}_{32}	a^{A}_{33}	a^{A}_{34}	0	a^{A}_{36}	0
#4,A Coal (as Finished Product) in A	0	0	0	a^{A}_{44}	0	0	0
#5,A Textiles in A	a^{A}_{51}	0	0	0	0	0	a^{A}_{57}
#1,B Labor in B	0	0	0	0	0	0	0
#2,B Transportation in B	0	0	0	0	$a^{A\to B}_{25}$	0	0
#3,B Coal (as Intermediate) in B	0	0	0	0	0	$a^{A\to B}_{36}$	0
#4,B Coal (as Finished Product) in B	0	0	0	0	0	0	0
#5,B Textiles in B	0	0	0	0	0	0	$a^{A\to B}_{57}$

TABLE 6. (cont'd).

COMMODITIES \ ACTIVITIES	#1,B Textile Production in B	#2,B Transportation Production in B	#3,B Coal (Intermediate) Production in B	#4,B Coal in B: Shift to Finished Product Status (Dummy)	#5,B Shipment of Transportation from B to A	#6,B Shipment of Coal (as Intermediate) from B to A	#7,B Shipment of Textiles from B to A
	(8)	(9)	(10)	(11)	(12)	(13)	(14)
#1,A Labor in A	0	0	0	0	0	0	0
#2,A Transportation in A	0	0	0	0	$a_{25}^{B\to A}$	0	0
#3,A Coal (as Intermediate) in A	0	0	0	0	0	$a_{36}^{B\to A}$	0
#4,A Coal (as Finished Product) in A	0	0	0	0	0	0	0
#5,A Textiles in A	0	0	0	0	0	0	$a_{57}^{B\to A}$
#1,B Labor in B	a_{11}^{B}	a_{12}^{B}	a_{13}^{B}	0	0	0	0
#2,B Transportation in B	a_{21}^{B}	a_{22}^{B}	a_{23}^{B}	0	a_{25}^{B}	a_{26}^{B}	a_{27}^{B}
#3,B Coal (as Intermediate) in B	a_{31}^{B}	a_{32}^{B}	a_{33}^{B}	a_{34}^{B}	0	a_{36}^{B}	0
#4,B Coal (as Finished Product) in B	0	0	0	a_{44}^{B}	0	0	0
#5,B Textiles in B	a_{51}^{B}	0	0	0	0	0	a_{57}^{B}

iated with commodities of region B are zero. This follows since for any production activity of region A a commodity of region B cannot be used an an input until it is shipped to region A and thus, as will be seen below, becomes transformed by another activity (shipment) into the corresponding commodity in region A, the region of production. In similar fashion, any output of a production activity in region A cannot be considered an intermediate or finished product in region B until it is shipped to region B, i.e. until another activity (shipment) takes place.

Columns 2 and 3 of Table 6, which refer to activities #2,A (transportation production in region A) and #3,A [coal (as intermediate) production in region A], do not introduce any new considerations. However, when we come to column 4, a new consideration is encountered. Whenever an item, such as coal in region A, can be used as both an intermediate and a finished product, and hence must be represented in our commodity framework by two commodities, such as coal (as intermediate) and coal (as finished product), we also must introduce into our framework another activity in addition to the commodity production activities. We let the activities which produce the commodity produce the commodity **as intermediate**; the additional activity then refers to the shift of the status of the given commodity from that of an intermediate to that of a finished product, i.e. to the transformation of the **intermediate** commodity into a **finished product** commodity. Thus, in Table 6, activity #3,A refers to the production of coal as intermediate in region A; and activity #4,A, in a sense a dummy activity, refers to shift of coal (as intermediate) to a finished product status, i.e. to the transformation of coal (as intermediate) in region A into coal$_A$ (as finished product) in region A. The corresponding coefficients are a^A_{34} (negative, representing an input) and a^A_{44} (positive, representing an output) where typically $-a^A_{34} = a^A_{44} = 1$.

Columns 5, 6 and 7 of Table 6 refer to shipment activities, i.e. to shipment of commodities in region A to region B. As noted above, such shipment effects a transformation of the commodity of region A (which may be taken to represent an input) into the corresponding commodity of region B (which represents an output). Thus column 6 which refers to activity #6,A, i.e. the shipment of coal (as intermediate) from A to B has as inputs the necessary transportation per unit level (i.e. a^A_{26})[42] and the coal in region A to be shipped per unit level (i.e. a^A_{36}) and as output coal in region B (i.e. $a^{A\to B}_{36}$). Note that the coefficient representing output of a shipment has associated with it two regional superscripts, the first the region of origin, and the second the region of termination. Such double superscripting is necessary in order to distinguish this output (i.e. $a^{A\to B}_{36}$) from the corresponding input a^A_{36} and from the input or output a^B_{36} associated with activity #6 of region B. The arrow between the two superscripts serves an obviously useful purpose.[43]

[42] Recall the convention that the transportation service required for an export is provided by the region of export.

[43] Since in this section double superscripts will be encountered only in connection with shipment activities, this notation facilitates the identification of shipment activities and corresponding exports and imports. In turn, the set of constraints is more easily constructed, and the basic structure of a particular problem more easily perceived.

Columns 5 and 7 follow the pattern of column 6 and require no additional comment. Note, however, that in column 5, which refers to the shipment of transportation services from region A to region B, the coefficient a^{A}_{25} covers not only the transportation service which is shipped (exported) from region A to region B but also the transportation service which is required to effect that shipment (export).

In similar manner, the activities of region B are described by columns 8-14 of Table 6. The coefficients of these columns bear the same interpretation as those already discussed. Observe that all the coefficients relating to commodities of region A are zero, except where a commodity is an output of a shipment activity of region B and thus enters region A as an import.[44]

Given the above activity coefficient matrix -- from which consumption activities and activities producing joint products have been excluded[45] -- we proceed to the statement of the objective function. Since we wish to maximize the sum of regional incomes, the objective function is:

$$\text{(18)} \qquad \begin{aligned} \text{Max } Y = {} & c^{A}_{1}x^{A}_{1} + c^{A}_{2}x^{A}_{2} + c^{A}_{3}x^{A}_{3} + c^{A}_{4}x^{A}_{4} + c^{A}_{5}x^{A}_{5} + c^{A}_{6}x^{A}_{6} + c^{A}_{7}x^{A}_{7} \\ & + c^{B}_{1}x^{B}_{1} + c^{B}_{2}x^{B}_{2} + c^{B}_{3}x^{B}_{3} + c^{B}_{4}x^{B}_{4} + c^{B}_{5}x^{B}_{5} + c^{B}_{6}x^{B}_{6} + c^{B}_{7}x^{B}_{7}. \end{aligned}$$

The X's are the choice variables, and the c's are the multiplying constants which represent income per unit level. As indicated above, the c for any given activity is obtained by multiplying that activity's unit level outputs of finished products by the **given** finished product prices. The given prices in our model are the price of coal$_A$ (as a final product) in region A ($\bar{P}^{A}_{4}$), the price of textiles in region A ($\bar{P}^{A}_{5}$), the price of coal (as finished product) in region B ($\bar{P}^{B}_{4}$) and the price of textiles in region B ($\bar{P}^{B}_{5}$). Note that as before we place a bar over each of these price symbols to indicate that each is a given constant. Thus for any c, say c^{B}_{4}, we have:

$$c^{B}_{4} = a^{B\text{->}A}_{44}\bar{P}^{A}_{4} + a^{B\text{->}A}_{54}\bar{P}^{A}_{5} + a^{B}_{44}\bar{P}^{B}_{4} + a^{B}_{54}\bar{P}^{B}_{5}.$$ [46]

[44] The activities of region B correspond in kind and number to those of region A. This symmetry is adopted to facilitate exposition. In the typical model, such symmetry would not appear. Certainly, in situations where one region possesses a valuable deposit of an immobile resource not present in a second region, such symmetry could not exist (except in a formal mathematical sense, where the cost of extracting the resource in the second region might be taken as infinite). In general, sets of activities different in both kind and number would confront the several regions of any interregional system.

Note also that our model assumes that the resource of each region, namely, labor is immobile. This assumption is not required. When a resource is mobile, we need to introduce into each region one additional shipment activity, namely the shipment of the resource from the respective region.

[45] The incorporation into the matrix of such activities will be discussed later.

[46] Commodities #1 to #3 in each region are resources or intermediates, upon the consumption of each of which a constraint will be imposed. Their prices are not **given**s of the problem, and will be derived (as shadow prices) from the solution of the problem and its dual.

Since all the coefficients of the above equation are zero save a^B_{44} (see col. II, Table 6), this equation reduces to:

$$c^B_4 = a^B_{44}\bar{p}^B_4,$$

i.e. income per unit level of the activity #4,B is equal to the unit level of output of the finished product coal in region B multiplied by its price. In this fashion we obtain:

(19)

$c^A_1 = a^A_{51}\bar{p}^A_5$	$c^B_1 = a^B_{51}\bar{p}^B_5$
$c^A_2 = 0$ (see fn. 47)	$c^B_2 = 0$ (see fn. 47)
$c^A_3 = 0$ (see fn. 47)	$c^B_3 = 0$ (see fn. 47)
$c^A_4 = a^A_{44}\bar{p}^A_4$	$c^B_4 = a^B_{44}\bar{p}^B_4$
$c^A_5 = 0$ (see fn. 47)	$c^B_5 = 0$ (see fn. 47)
$c^A_6 = 0$ (see fn. 47)	$c^B_6 = 0$ (see fn. 47)
$c^A_7 = a^A_{57}\bar{p}^A_5 + a^{A\to B}_{57}\bar{p}^B_5$	$c^B_7 = a^{B\to A}_{57}\bar{p}^A_5 + a^B_{57}\bar{p}^B_5$

The particular significance of each of these constants will be discussed below in connection with the presentation of the corresponding constraint in the dual.

Since in our simple linear programming problem, we have in each region one resource and two intermediate commodities, we have three constraints in each region. They are listed in the Part I of Table 7 under the objective function. The first constraint states that the total requirement of commodity #1,A (the labor resource in region A) by all activities must not exceed the available stock of this commodity. Similarly, the fourth constraint states that the requirement of commodity #1,B (the labor resource in region B) must not exceed the available stock of this commodity.

The second, third, fifth and sixth constraints refer to intermediate commodities. They therefore read differently. For example, the second constraint reads horizontally, term by term: (1) the requirement of transportation by the textile production activity in region A $(-a^A_{21}x^A_1)$ [48] **minus** (2) the output of transportation by the transportation production activity in region A $(a^A_{22}x^A_2)$ **plus** (3) the requirement of transportation by the coal production activity in region A $(-a^A_{23}x^A_3)$ **plus** (4) the exports of transportation from region A as well as the requirement of transportation

[47] Activities #2, #3, #5 and #6 in both regions produce intermediates only. Hence, as is to be expected from the discussion of section 7, income per unit level of these activities is zero.

[48] Recall that the coefficient a^A_{21} is negative if not zero.

to effect this export $(-a^{A}_{25}x^{A}_{2})$,[49] i.e. the transportation demand of the transportation shipping activity of region A **plus** (5) the requirement of transportation by the coal shipping activity of region A $(-a^{A}_{26}x^{A}_{6})$ **plus** (6) the requirement of transportation by the textile shipping activity of region A, and **minus** (7) the import of transportation by region A $(a^{B->A}_{25}x^{B}_{5})$ which is the output of the transportation shipping activity of region B -- all these seven items must yield a result not greater than zero. Or, as stated previously, the demand for transportation by local activities plus exports must not exceed local output plus imports.

In a similar manner, we may interpret the third, fifth and sixth constraints. To complete the statement of our interregional linear program, we add the requirement that each activity level be non-negative, as indicated in the last row of Part I of Table 7.

The significance of our interregional linear program can be more clearly discerned if the dual is presented. This is done in Part II of Table 7. In the first row of Part II we state the objective function, which is to select non-negative prices for the two resources, labor in region A and labor in region B, to minimize W, i.e. fictitious returns to resource owners. Note that in the objective function there are six terms, each involving a variable price.[50] In each term a variable price (which is unbarred to indicate that it is a variable) multiplies the corresponding R (the limiting constant) in the constraints of the linear program. But four of these R's (specifically the R's which limit consumption of **intermediate** commodities) have a value of zero. Hence, whatever the level of the corresponding P's (which represent the prices of the corresponding intermediate commodities), the value of the objective function is unaffected.[51]

Since there are as many constraints in a dual as there are choice variables in the original linear program, we have fourteen constraints in our dual. They are listed in Part II of Table 7. The limiting constant (i.e. the **c**) for each constraint is the corresponding constant which multiplies the choice variable in the linear program. Its value, which is the income per unit level of the corresponding activity, has already been determined along lines discussed above. Observe, too, that the set of coefficients contained in the constraints of the dual is obtained by interchanging the rows and columns of the set of coefficients contained in the linear program. Thus the first column of a's in the constraints of the linear program becomes the first row of a's in the constraints of the dual, etc.

Each constraint of the dual states that for any given activity the value of inputs of resources and intermediates less the value of the output

[49] Recall that $-a^{A}_{25}$ covers per unit level the transportation to be exported as well as the transportation input to effect this export.

[50] Thus there are as many choice variables in the dual as constraints in the original linear program.

[51] Actually, however, for any feasible set of prices for labor in both regions, the prices of the intermediate commodities are determined by the constraints of the dual to be discussed below.

TABLE 7. Part I: An Interregional Linear Program.

| | | |

TABLE 7. Part II: The Dual.

Min: $W = P_1^A R_1^A + P_2^A \cdot 0 + P_3^A \cdot 0 + P_1^B R_1^B + P_2^B \cdot 0 + P_3^B \cdot 0$													
subject to:													
1.	$-a_{11}^A P_1^A$	$-a_{21}^A P_2^A$	$-a_{31}^A P_3^A$									$\geq$	c_1^A
2.	$-a_{12}^A P_1^A$	$-a_{22}^A P_2^A$	$-a_{32}^A P_3^A$									$\geq$	$c_2^A = 0$
3.	$-a_{13}^A P_1^A$	$-a_{23}^A P_2^A$	$-a_{33}^A P_3^A$									$\geq$	$c_3^A = 0$
4.			$-a_{34}^A P_3^A$									$\geq$	c_4^A
5.		$-a_{25}^A P_2^A$				$-a_{25}^{A\to B} P_2^B$						$\geq$	$c_5^A = 0$
6.		$-a_{26}^A P_2^A$	$-a_{36}^A P_3^A$						$-a_{36}^{A\to B} P_3^B$			$\geq$	$c_6^A = 0$
7.		$-a_{27}^A P_2^A$										$\geq$	c_7^A
8.				$-a_{11}^B P_1^B$	$-a_{21}^B P_2^B$	$-a_{31}^B P_3^B$						$\geq$	c_1^B
9.				$-a_{12}^B P_1^B$	$-a_{22}^B P_2^B$	$-a_{32}^B P_3^B$						$\geq$	$c_2^B = 0$
10.				$-a_{13}^B P_1^B$	$-a_{23}^B P_2^B$	$-a_{33}^B P_3^B$						$\geq$	$c_3^B = 0$
11.									$-a_{34}^B P_3^B$			$\geq$	c_4^B
12.		$-a_{25}^{B\to A} P_2^A$				$-a_{25}^B P_2^B$						$\geq$	$c_5^B = 0$
13.			$-a_{36}^{B\to A} P_3^A$			$-a_{26}^B P_2^B$			$-a_{36}^B P_3^B$			$\geq$	$c_6^B = 0$
14.						$-a_{27}^B P_2^B$						$\geq$	c_7^B
$P_1^A \geq 0$; $P_2^A \geq 0$; $P_3^A \geq 0$; $P_1^B \geq 0$; $P_2^B \geq 0$; $P_3^B \geq 0$													

of intermediates, if there is any such output, must not be less than the income per unit level of that activity. Thus, the first constraint states that per unit level of textile production in region A the cost of labor inputs ($-a^A_{11}P^A_1$) plus the cost of transport inputs ($-a^A_{21}P^A_2$) plus the cost of coal inputs ($-a^A_{31}P^A_3$) must not be less than income per unit level (c^A_1), where, as above, this income per unit level is equal to sales value of the unit level output of textiles ($a^A_{51}P^A_5$). Thus, too, the second constraint states that per unit level production of the intermediate transportation the cost of labor inputs ($-a^A_{12}P^A_1$) minus the value of transportation output ($a^A_{22}P^A_2$) plus the cost of coal inputs ($-a^A_{32}P^A_3$) must not be less than income per unit level of this intermediate production activity (c^A_2) which is always zero since no finished product commodities are produced.[52]

Another type of constraint is the fourth. It simply states that per unit level of the dummy activity #4,A which involves the shift of coal from an intermediate to a finished product status in region A, the cost of the intermediate coal ($-a^A_{34}P^A_3$) must not be less than the value of the output of coal as a finished product (c^A_4 which equals $a^A_{44}P^A_4$). If we set $-a^A_{34} = a^A_{44} = 1$, we have that in region A the price of the intermediate coal must not be less than the price of the finished product coal. When the dummy activity #4,A is operated at a positive level in an optimal solution of the linear program, this implies that the price of coal as an intermediate does not exceed, and hence equals the given price of the finished product coal.

Still another type of constraint derives from the shipping activities of the linear program and interrelates prices among regions. For example, take the sixth constraint. It states that per unit level of activity #6,A, which ships coal (as intermediate) from region A to region B, the cost of transport ($-a^A_{26}P^A_2$) plus the value of the coal to be exported ($-a^A_{36}P^A_3$) less the value of coal imports (deliveries) in the region of import ($a^{A->B}_{36}P^B_3$) must not be less than zero ($c^A_6 = 0$). Or, if we let $-a^A_{36} = a^{A->B}_{36} = 1$, we have the price of coal (as intermediate) in the region of export (region A) plus the cost of transport must not be less than the price of coal (as intermediate) in the region of import (region B). When the operation of the coal shipping activity is positive in the optimal solution to the linear program, such operation implies that the price of the intermediate coal in the region of import does equal the price of the intermediate coal in the region of export plus transport cost. This equality is the familiar equilibrium relation obtained when several regions trade under conditions of perfect competition.

The seventh constraint of the dual is a variant of the sixth. It states that per unit level of activity #7,A which refers to the export of textiles from region A to region B, the transportation cost of the textiles to be exported ($-a^A_{27}P^A_2$) must not be less than unit level income (c^A_7 where

[52] As already indicated, when the intermediate transportation is produced and when a unit level of the transportation producing activity is defined so that a^A_{22} is unity, this relation implies that the price of transportation equals unit cost, i.e. the cost of the required labor and coal per unit output.

$c_7^A = a_{57}^A \bar{P}_5^A + a_{57}^{A->B} \bar{P}_5^B$). In effect this seventh constraint states that where $-a_{57}^A = a_{57}^{A->B} = 1$ the price spread among the two regions must not exceed the cost of transport.[53] When the operation of this activity is positive in the optimal solution to the linear program, this operation implies that the price spread just equals the transport cost. Such a result is once again consistent with equilibrium conditions for trade among regions under conditions of perfect competition.

9. THE CHOICE OF FINISHED PRODUCT PRICES

The last paragraphs of the above section treat the several interpretations to be given to the fourteen constraints contained in the dual. However, in connection with the last constraint discussed, a major problem arises. As already indicated several times, the linear program requires that the set of final product prices be predetermined, i.e., given beforehand. More specifically in our problem, the prices of the finished products, textiles and coal, must be specified at the start. But this would seem to imply a previous knowledge of what the solution to the linear program would be. To illustrate, consider the price of textiles in both regions. If beforehand we set a price for the finished product textiles in region B which is greater than the price of the finished product textiles in region A, we thereby have predetermined that the level of textile shipment from region B to region A, namely, activity #7,B, be zero. Or if at the start we set the price of textiles in region A higher than in region B, this step precludes the shipment of textiles from region A to region B, i.e. forces activity #7,A to run at zero level. Or if we set the two prices equal to each other, this step precludes any shipment in either direction, i.e. specifies that activities #7,A and #7,B be run at zero level (unless the price of transport service were to reach the extreme of zero).

The above discussion clearly demonstrates that for our interregional linear programming model a consistent and meaningful set of final product prices must be established at the start. The necessity for such a step is the more evident in an interregional system composed of many regions and many commodities, such as the generalized system to be presented below. The problem is simply this: if the technique is to determine among others the levels of shipping activities of an optimal program, then one must not initially hypothesize finished product prices which preclude such determination. Yet at the same time, one cannot carry through the linear programming computation without setting these prices.

The above dilemma is more apparent than real. In constructing any linear program, the researcher necessarily must have in the background certain hypotheses about feasible, likely, or desirable consumption levels. He will therefore set finished product prices which are not inconsistent with these levels.[54] For example, if wheat is to be consumed both in

[53] Recall that a_{57}^A is negative since it indicates an input. Hence if $-a_{57}^A = 1$, then $a_{57}^A = -1$ and $c_7^A = \bar{P}_5^B - \bar{P}_5^A$, i.e. the price spread.

[54] In certain cases, such levels and prices will be related to levels and prices currently existing in an interregional economy.

Philadelphia and Pittsburgh and if the pattern of land resources dictates that wheat can be grown productively only to the west of these cities, then he must not set the price of wheat at Pittsburgh higher than at Philadelphia. Nor may the analyst set the price of wheat at zero at Pittsburgh, or hypothesize an unreasonable spread of prices between Philadelphia and Pittsburgh. In general, he will select a set of prices which he judges will differ by an amount equal to the difference between transport costs on a unit of wheat to Philadelphia and to Pittsburgh. However, he bears in mind that the price of transport service, and thus transport costs on wheat, is to be determined from the linear programming computation; hence he can at best make only a rough approximation of a meaningful spread of prices.

Despite all the ingenuity that the researcher may apply, he is most likely to find that an optimal program schedules the shipment of all wheat either to Pittsburgh, Philadelphia, or one or several other locations. This result would reflect the fact that his estimate of the difference in transport cost to these two locations, or other locations was incorrect, even if by only a small amount. Thus in the problem such as we have posed, the analyst must **either** (1) introduce additional constraints which guarantee that the amount of each finished product available for household consumption in each region shall achieve at least a certain level (although he must be careful not to stipulate amounts which cannot be met by the program), or (2) introduce another set of activities, namely, consumption activities. Because of limitations of space we shall develop below only the first of these alternatives.

Consider the introduction of additional constraints in each region on the amounts of the two finished products, textiles and coal, to be made available for household consumption.[55] Suppose we state that this amount of the finished product coal in region A shall be at least equal to the quantity R_4^A, of textiles in region A at least equal to R_5^A, of the finished product coal in region B at least equal to R_4^B, and of textiles in region B at least equal to R_5^B. We are therefore stating for region A that:

(1) the amount of the finished product coal yielded by activity #4,A ($a_{44}^A x_4^A$) must be at least as great as the required amount of coal to be made available for household consumption in region A (R_4^A), i.e.

$$a_{44}^A x_4^A \geq R_4^A; \text{ and}$$

(2) the amount of the finished product textile yielded by activity #1,A ($a_{51}^A x_1^A$), less exports of textiles from region A to region B ($-a_{57}^A x_7^A$, which represent inputs of the textile shipping activity of region A), plus imports of textiles into region A from region B ($a_{57}^{B\to A} x_7^B$, which represent outputs of the textile shipping activity of region B) must be at least as great as the required amount of textiles to be made available for household consumption in region A (R_5^A), i.e.

$$a_{51}^A x_1^A + a_{57}^A x_7^A + a_{57}^{B\to A} x_7^B \geq R_5^A.$$

[55] The materials on consumption constraints have been adapted from Stevens, **op. cit.**, and other manuscript by the same author.

Similarly, for region B, we have the two comparable constraints:

$$a_{44}^{B} x_4^{B} \geq R_4^{B}; \text{ and}$$

$$a_{57}^{A->B} x_7^{A} + a_{51}^{B} x_1^{B} + a_{57}^{B} x_7^{B} \geq R_5^{B}$$

These four new constraints, which we shall call **availability constraints**, are to be added to the six specified in Part I of Table 7 to form the set of constraints applicable to the new linear program. However, since the linear program is a maximum problem, the sense of the inequality of each constraint must be "less than" (<). Therefore, to conform to this requirement, both sides of the above constraints are multiplied by -1, which then changes the sense of the inequality. In strict fashion the new constraints are:

$$\begin{aligned} -a_{44}^{A} x_4^{A} &\leq -R_4^{A} \\ -a_{51}^{A} x_1^{A} - a_{57}^{A} x_7^{A} - a_{57}^{B->A} x_7^{B} &\leq -R_5^{A} \\ -a_{44}^{B} x_4^{B} &\leq -R_4^{B} \\ -a_{57}^{A->B} x_7^{A} - a_{51}^{B} x_1^{B} - a_{57}^{B} x_7^{B} &\leq -R_5^{B} \end{aligned} \qquad (20)$$

Despite these new constraints, the objective function remains unchanged in form. However, the optimal solution for this new linear program will most likely yield a lower value for Y, the sum of regional incomes, if any of the **availability** constraints is binding. For, if one or more availability constraints are binding, this usually signifies that the system is being forced to yield a solution which would not be optimal for a system unconstrained by minimum limits on the amounts of finished products to be made available for household consumption in the several regions.[56] That is, in order to fulfill the requirement that a certain amount of a commodity be available for household consumption in a region, the system is being forced to ship a commodity to a region or produce a commodity in a region or curtail a region's export of a commodity, or some combination of these activities, in a way which the system would usually find not efficient.

Once availability constraints are added to a linear program, both the objective function and the set of constraints of its dual are changed. Because four new constraints are added to the linear program, we must add four new choice variables to the objective function of the dual, more specifically, $\tilde{p}_4^{A}$, $\tilde{p}_5^{A}$, $\tilde{p}_4^{B}$, and $\tilde{p}_5^{B}$.[57] These four variables, which can take

[56] Stevens, **op. cit.**, has designated this type of solution as "semi-efficient."

[57] The ~ symbol above the p's indicates that these prices are **variables**. They contrast with the corresponding p's representing finished product prices which are barred and are constants (fixed beforehand).

only non-negative values and which we shall find represent subsidy prices, multiply the corresponding new -R's which are predetermined constants. Thus, to the objective function at the top of Part II of Table 7, we add the following four terms:

$$+ \overset{''}{P}{}^{A}_{4}\,(-R^{A}_{4}) + \overset{''}{P}{}^{A}_{5}\,(-R^{A}_{5}) + \overset{''}{P}{}^{B}_{4}\,(-R^{B}_{4}) + \overset{''}{P}{}^{B}_{5}\,(-R^{B}_{5}), \text{ or}$$

$$- \overset{''}{P}{}^{A}_{4}R^{A}_{4} - \overset{''}{P}{}^{A}_{5}R^{A}_{5} - \overset{''}{P}{}^{B}_{4}R^{B}_{4} - \overset{''}{P}{}^{B}_{5}R^{B}_{5}$$

The number of constraints in the dual is not changed by the injection of availability constraints into the linear program. However, the statement of several of the constraints is changed by the addition to each of another term. Specifically constraints 1, 4, 7, 8, 11 and 14 become respectively:

1a $$-a^{A}_{11}P^{A}_{1} - a^{A}_{21}P^{A}_{2} - a^{A}_{31}P^{A}_{3} \qquad -a^{A}_{51}\overset{''}{P}{}^{A}_{5} \geq c^{A}_{1}$$

4a $$- a^{A}_{34}P^{A}_{3} - a^{A}_{44}\overset{''}{P}{}^{A}_{4} \qquad \geq c^{A}_{4}$$

7a $$- a^{A}_{27}P^{A}_{2} \qquad -a^{A}_{57}\overset{''}{P}{}^{A}_{5} - a^{A\to B}_{57}\overset{''}{P}{}^{B}_{5} \geq c^{A}_{7}$$

8a $$-a^{B}_{11}P^{B}_{1} - a^{B}_{21}P^{B}_{2} - a^{B}_{31}P^{B}_{3} \qquad -a^{B}_{51}\overset{''}{P}{}^{B}_{5} \geq c^{B}_{1}$$

11a $$- a^{B}_{34}P^{B}_{3} - a^{B}_{44}\overset{''}{P}{}^{B}_{4} \qquad \geq c^{B}_{4}$$

14a $$-a^{B}_{27}P^{B}_{2} \qquad -a^{B}_{57}\overset{''}{P}{}^{B}_{5} - a^{B\to A}_{57}\overset{''}{P}{}^{A}_{5} \geq c^{B}_{7}$$

We now wish to interpret these altered constraints. Actually, we need to examine carefully only the first three of these constraints (which pertain to activities in region A) since the last three (which pertain to activities in region B) are identical to the first three except that regional superscripts have been interchanged.

To begin, rearrange terms so that all $\overset{''}{P}$'s are on the right hand side, and substitute for c^{A}_{1}, c^{A}_{4} and c^{A}_{7} their definition in terms of finished product prices, as given above in the set of equations (19). Assuming that we have defined activities #1,A, #4,A and #7,A so that at unit level of operations each yields a unit of product, i.e. that respectively $a^{A}_{51} = 1$, $a^{A}_{44} = 1$ and $a^{A\to B}_{57} = -a^{A}_{57} = 1$, we obtain for the first three of the altered constraints:

1b $$-a^{A}_{11}P^{A}_{1} - a^{A}_{21}P^{A}_{2} - a^{A}_{31}P^{A}_{3} \geq \bar{P}^{A}_{5} + \overset{''}{P}{}^{A}_{5}$$

4b $$- a^{A}_{34}P^{A}_{3} \geq \bar{P}^{A}_{4} + \overset{''}{P}{}^{A}_{4}$$

7b $$- a^{A}_{27}P^{A}_{2} \qquad \geq (\bar{P}^{B}_{5} + \overset{''}{P}{}^{B}_{5}) - (\bar{P}^{A}_{5} + \overset{''}{P}{}^{A}_{5})$$

Constraint 1b states that per unit level of textile production in region A, the cost of labor ($-a^A_{11}P^A_1$) plus the cost of transportation ($-a^A_{21}P^A_2$) plus the cost of coal ($-a^A_{31}P^A_3$) must not be less than the predetermined price of textiles in region A ($\bar{P}^A_5$) plus the value attached to the availability (for household consumption) of a unit of textiles in region A (P^A_5). This latter value, P''^A_5, may be designated a subsidy price. In one sense, it is the opposite of a rent. Rent on a resource is zero when the demand for a resource is less than the **fixed** available supply. When the demand would exceed the **fixed** supply if the resource were costless, rent becomes positive, thereby serving to curtail demand to a level equal to the fixed supply. In contrast, when the amount of textiles made available to region A exceeds the fixed availability requirement, i.e. R^A_5 (which some may wish to view as a minimum acceptable level of household consumption), P''^A_5 is zero. However, when the amount of textiles made available to region A would be less than the constant R^A_5 if P''^A_5 were zero, then P''^A_5 becomes positive. It represents a subsidy price[58] to be added to the specified market price in order to raise this amount (supply) to a level equal to R^A_5. Thus the variable P^A_5 serves to achieve equality of supply and demand by affecting the behavior of suppliers (producers and shippers); whereas a rent variable, say P^A_3, serves to achieve equality by affecting the behavior of demanders.

In similar fashion, constraint 4b states that per unit level of activity #3,A the cost of the input of coal as an intermediate (which is the price of coal as an intermediate since $-a^A_{34} = 1$) must not be less than the price of the finished product coal in region A plus the subsidy price on the finished product coal in region A. This subsidy price will be zero when the corresponding availability constraint is not binding; otherwise, it will usually be positive.

Finally, constraint 7b states that per unit level of activity #7,A the cost of inputs, which is the transport cost on a unit of textiles from region A to region B, must not be less than the difference between (a) the price of textiles plus subsidy price on textiles in region B and (b) the price on textiles plus subsidy price on textiles in region A. In cases where the system furnishes to the respective region supplies of textiles which exceed R^A_5 and R^B_5, both subsidy prices are zero. In other cases, where only one of the regions receives supplies of textiles in excess of its availability requirement, the corresponding subsidy price is zero, while the subsidy price of the second region is usually positive since the second region's availability constraint on textiles is binding. In still a third set of cases, the availability constraints of both regions are binding, and both subsidy prices are usually positive. In each case of this third set, the operation of the system, **if the system were free of availability constraints**, would furnish the respective regions with supplies of textiles short of R^A_5 and R^B_5. Injecting availability constraints into the system forces, for these cases the contraction of the output of coal as a finished product in one or both regions and diverts resources to the provision of additional textiles in both regions via the mechanism of subsidy prices on textiles.

[58] Alternatively, P''^A_5 might be considered a location utility price, or a premium price depending upon the manner and situation in which a linear program is conceived.

In the above manner, one overcomes the problem of setting finished product prices which are required for the operation of the model and yet not predetermine levels of activities. By the introduction of availability constraints, we bypass the dilemma posed above. At the same time we derive a set of subsidy prices which when added to the preassigned finished product prices gives us a set of meaningful prices which we shall designate **adjusted finished product prices.** The adjusted finished product prices are variables, since each is the sum of a constant (the preassigned finished product price) plus a variable adjustment (the variable subsidy price). They are meaningful because we can now operate the model without the availability constraints, but with the adjusted finished product prices as determined above substituted for the first set of preassigned finished product prices (the $\bar{P}$'s); we can do this and yet be certain that in each region supplies of each finished product made available will at least equal the availability requirements (i.e. the desired levels). The adjusted finished product prices are therefore a meaningful and consistent set of prices for an interregional system.

Another point is to be noted. Suppose an analyst operates a model not restricted by availability constraints with the adjusted finished product prices as given data. Suppose he finds that the resulting regional distribution of finished product available for household consumption is not consistent with **disposable** regional incomes (however these incomes are determined). He can then introduce once again availability constraints. On any finished product he can set an availability requirement in, let us say, a high income region which exceeds the supply available for household consumption, while he sets an availability requirement in, let us say, a low income region which falls short of supply available for household consumption. With such new availability constraints on one or more commodities for one or more regions, and with the derived adjusted finished product prices as given data, he can operate the resulting new linear program. This new linear program will force shifts in the regional distribution of available supplies of finished products in accordance with the judgment of the analyst, as reflected in the new set of availability requirements. The dual of this new linear program will yield a set of subsidy prices which when added to the first set of derived adjusted finished product prices yields a second set of adjusted finished product prices which are meaningful and consistent in terms of the new availability requirements imposed. Again, the investigator may compare the resulting regional pattern of available supplies with the resulting new set of **disposable** regional incomes (however determined), and reach the decision to alter availability requirement by regions once again, etc., etc. By this iterative procedure, the investigator can work toward a final set of interregional prices consistent with his other research hypotheses, data and findings, and toward a system which optimizes the sum of regional incomes within this latter framework.

10. THE GENERAL INTERREGIONAL MODEL

We are now in a position to generalize. Let there be U regions, any one region generally being designated by the superscript L, L = 1,...,U.[59]

[59] In a pure theoretical model there is of course no restriction on the number of regions to be considered, provided that the number if finite. Hence, in con-

Let there be in each region many resources. A resource may be either mobile or immobile; its initial supply (endowment) in any given region is fixed and specified beforehand. Also, let there be in each region many commodities, some of which are pure intermediates, some of which are pure finished products, and some of which can be consumed either as intermediates (by industry) or as finished products (by households). Moreover, let there be a commodity classification which, for ease of exposition alone, we take to be identical from region to region. In this classification, commodities are designated from 1 to s. Resources and intermediate commodities run from 1 to m, commodity i being generally taken to represent a resource or intermediate. Finished products run from (m+1) to s, commodity q being generally taken to represent a finished product. As before, any item which is both an intermediate and a finished product is considered as two commodities; it is listed once as **intermediate** among the first m commodities; and it is listed a second time as **finished product** among commodities designated from (m+1) to s. Moreover, a resource, such as labor which is used as an input by industry but also may be consumed by households as **leisure**, may appear twice in the commodity classification, once among the first group of commodities as an **industrial resource** and again among the second group of commodities as **finished product**. For notational convenience, we order the finished products such that items which appear only as finished products run from (m+1) to r, and items which appear both as finished product and as an intermediate (or resource) run in the finished product classification from (r+1) to s.

In general we let R_i^L $(i = 1,\ldots,m;\ L = 1,\ldots,U)$ represent the initial supply of resource or intermediate i in region L. This initial supply is fixed and specified beforehand. In the case of an intermediate, $R_i^L = 0$.

Associated with each region is a set of activities. Some of these are production activities; others are dummy activities which shift intermediates and resources to a finished product status; and still others are shipping activities. Again, for ease of exposition alone, we assume that the set of activities is identical from region to region, although this will not be the case for any given interregional situation.

---

ceptually approximating reality, one can envisage as many **one-point** regions as he desires. One can always expand the number of regions if he considers that the assumed number does not adequately represent the existing clusters of population, economic and other units.

It is important to recognize that in practice the number of regions which is adopted depends upon the existing interregional situation to be analyzed. This dependence results from the structure of a linear programming model which necessitates that constant coefficients characterize the operation of each activity, including shipping. Thus, the transport inputs per unit level of any shipping activity must be constant. This in turn necessitates that each region should be capable of being approximately represented by a single-point economy where at are concentrated all production and consumption activities of the region. Typically, too, as in this model, the assumption is made that intraregional requirements of transport inputs are zero. This last assumption is convenient, but not necessary, since one can incorporate into his scheme intraregional shipping activities. Each of these shipping activities, however, must require a constant amount of transport inputs per unit level; i.e. the use of transport inputs in shipping any given commodity within a region must be approximately a constant regardless of point of origin and point of destination within the region.

For each of the U regions we assume n production activities. As activities #1,A and #2,A in Table 6, each production activity employs as inputs resources and intermediate commodities. However, unlike the production activities of Table 6, each production activity may yield more than one output. It may yield more than one finished product, or more than one intermediate, or some combination of intermediates and finished products. Thus, more than one of the coefficients listed in the appropriate column of the coefficient matrix for this general interregional model may be positive.[60] Further the same finished product, or intermediate, or combination of finished products and intermediates may be produced by more than one activity, i.e. there can be in each region activities (processes) producing the same kind of output.

For each region, too, we have the same number of dummy activities, each one of which transforms an intermediate, or in some cases a resource, into a finished product. As with activities #4,A and #4,B in Table 6, each dummy activity has a single input and a single output, both of the same magnitude but opposite in sign. Since in each region there are m resources and intermediates, we may posit as many as m dummy activities in each region.[61]

Finally, for each region we have the same number of shipping activities. Any region L may ship any one commodity to (U-1) regions. Since region L is generally at a different distance from each of these (U-1) regions, its requirements of transport inputs to ship the commodity to each of these regions will be different. Hence, there must be (U-1) shipping ac-

[60] Introduction of joint products for any production activity does not involve any complication. Of course, where more than one finished product is associated with a production activity, income per unit level of that activity will be the sum of more than one term, each term representing the product of price of a finished product and its output per unit level.

It is to be noted that when activities are defined such that an activity both produces the output and ships it to a specified region (i.e. requires as inputs the commodities normal to the production process as well as transport services for delivery), then difficulty arises when more than one product is associated with a productive process. For suppose two finished products are yielded by a productive process. If both are delivered to region J, a defined amount of transport inputs are required per unit level of the activity embodying this productive process and entailing delivery of the two finished products to J. However, if one product is shipped to region J and a second to region L, then another defined amount of transport inputs will be required per unit level, and a second activity will need to be recognized. Further, if one product is shipped to J, and the second to K; or if one is shipped to J, and one-half the second shipped to both L and K; etc., a third, fourth, etc. activity will need to be recognized, since each delivery pattern will require a different amount of transport inputs per unit level. Since an infinite number of delivery patterns is possible when more than one output is yielded by a productive process, an infinite number of shipping activities would need to be recognized. This would negate the usefulness of the model. Hence, if joint products are permitted in a production process, it becomes necessary to divorce production activity from shipping activity as we have done in our model.

[61] In the case of a dummy activity which converts a resource, say labor, into a finished product, say leisure, we have the single input, labor, and the single output, leisure.

tivities in region L to allow for export of the given commodity to each region. But there are as many as r commodities in region L which may be shipped.[62] As a consequence there are (U-1)r shipping activities in region L. Each shipping activity has associated with it a single output, the commodity which is received as an import in the region of destination. Except for activities which export transportation, such as activities #5,A and #5,B in Table 6, each shipping activity requires at least two inputs, one being the commodity to be exported, the other being transportation. Generally speaking, each shipping activity is defined to require other inputs as well, such as management, insurance services, etc.

Altogether there are in region L, as well as every other region, n+m + (U-1)r possible activities. For notational convenience we let v represent the number of these activities, i.e. we set v = n+m + (U-1)r. We now number these activities from 1 to v, the first n being production activities, the next m being dummy activities, and the remaining (U-1)r being shipping activities. For any activity j in region L, we let X^L_j represent the level at which that activity is operated. Since there are v activities in each region (i.e. j = 1,...,v) and since there are U regions (i.e. L = 1,...,U), there are all told Uv activities in our interregional system.

Corresponding to every activity in every region there is an income per unit level. That is, corresponding to any activity j in region L, there is a c^L_j. As in the discussion of sections 7 and 8, c^L_j is defined as the sum of the products obtained by multiplying each finished product output of an activity by the given price of the finished product, i.e. for all non-shipping activities:

$$(21) \qquad c^L_j = a^L_{m+1,j}\bar{p}^L_{m+1} + a^L_{m+2,j}\bar{p}^L_{m+2} + \ldots + a_{sj}\bar{p}^L_s = \sum_{q=m+1}^{s} a^L_{q,j}\bar{p}^L_q;$$

and for all shipping activities, each of which by definition ships only a single commodity:

$$(22) \qquad c^L_j = 0 \text{ when an intermediate is shipped; or}$$

$$(23) \qquad c^L_j = a^L_{qj}\bar{p}^L_q + a^{L\to J}_{qj}\bar{p}^J_q \qquad \text{when a finished product is shipped.}[63]$$

[62] Finished products which run from (r+1) to s are not shipped since these items have as a counterpart an intermediate (or resource) and since we adopt the convention that any item which is both a finished product and an intermediate (resource) is shipped in its intermediate (resource) form. The use of another convention at this point would not basically alter the ensuing analysis.

[63] It is to be borne in mind that the c's change when the set of predetermined prices on finished products changes. Thus, if after a first run of the model, the resulting regional distribution of finished product available for household consumption is not consistent with disposable regional incomes, the investigator may try another set of finished product prices -- one that is adjusted in the light of the resulting set of subsidy prices. He may place new prices upon such finished products as textiles and leisure and perhaps introduce new availability constraints relating to these finished products. In doing so, the c's change at the same time.

If the aim of the linear program is to maximize the sum of regional incomes, the objective function is:

$$\text{(24)} \qquad \text{Max } Z = c_1^1 x_1^1 + \ldots + c_v^1 x_v^1 + c_1^2 x_1^2 + \ldots + c_j^L x_j^L + \ldots + c_v^U x_v^U = \sum_{L=1}^{U} \sum_{j=1}^{v} c_j^L x_j^L$$

for non-negative X_j^L. In this program, there may be as many as Us constraints, which include: (a) constraints on the use of resources,[64] immobile or mobile; (b) constraints on intermediates; and (c) availability constraints on finished products. There may be as many as Um of the first two types, and as many as U(s-m) of the third type. The general form of these constraints is:

$$\text{(25)} \qquad -\sum_j a_{kj}^L x_j^L - \sum_{J \neq L} \sum_j a_{kj}^{J \to L} x_j^J \leq R_k^L \qquad \begin{matrix} k=1,\ldots,s \\ L=1,\ldots,U \end{matrix}$$

If k is an immobile resource in region L, the second term of this constraint is zero since all $a_{kj}^{J \to L}$ are zero (no shipment possible). Thus, only the first term remains which represents the sum of the requirements of this resource by all production activities in region L, by a possible dummy activity in region L which transforms this resource into a finished product, and by all shipping activities in region L where these shipping activities require the resource as local inputs and not for export. This sum must not be greater than R_k^L, the available supply in region L.

If k is a mobile resource, to the above sum must be added the requirements of this resource by shipping activities for export to other regions. Also, the second term may no longer be zero. In this term, for all but one j in any other region, $a_{kj}^{J \to L} = 0$, since, as noted in section 8, no production, dummy or shipping activity in region L can use as inputs commodities of other regions which have not first been shipped to region L and thus transformed into the corresponding local commodity in region L. The one activity, however, for which $a_{kj}^{J \to L} \neq 0$ is that shipping activity which transports the resource from region J to region L.[65] Thus the absolute value of the second term of relation (25) represents region L's total im-

[64] Resources are defined very broadly so as to include not only specific types of natural resources, such as land of a certain soil-type, or ore of a certain quality, or coking coal of a certain grade, but also non-natural resources such as industrial plant of a certain type, equipment of a defined specification, or labor of a certain skill, etc. Limited supplies (capacities) of the latter type of resources are of particular importance in short-run models, such as some of the input-output schemes.

[65] Since the position of this resource shipping activity in the listing of activities will vary from region to region, we cannot designate by a specific subscript that activity in each region J ($J \neq L$) which ships this resource to region L.

ports of the resource. Thus, for such a resource, relation (25) states that the requirements of this resource as local inputs by all activities plus requirements for export less total imports must not be greater than the given local supply.

If k in (25) represents an intermediate, then the constraint reads: the requirements of k by all production activities in region L, plus the requirement of k by the dummy activity of region L which transforms k into a finished product, plus the requirement for both local use and export by region L's shipping activities, less the production of k by activities in region L, and less the imports of k from other regions [the absolute value of the second term of (25)][66] must not be greater than zero, which is the value R_k^L takes when k is an intermediate. Or, in other words, the demand for the intermediate k must not exceed its supply.

Finally, if k is a finished product [i.e. runs from (m+1)...s], relation (25) states the following, **after we multiply both sides by -1 and change the sense of the inequality, which we may do without altering the relation:** The sum of the outputs of the finished product k by all production and dummy activities of region L less the amount taken by shipping activities for export to other regions plus imports of k from all other regions (i.e. the outputs of those shipping activities in the several regions which ship k to region L) must not be less than $-R_k^L$ which is a positive amount fixing the quantity of the finished product k which must be available for household consumption in region L. Note a change in the definition of R_k^L. In the previous section, R_k^L was defined as a non-negative number. Here we define R_k^L as a non-positive number (but of the same absolute value) so that $-R_k^L$ is non-negative.[67] We adopt this change in notation in order to embrace all constraints of the problem in a single statement of relations (25).

Observe that there is specified in this general model an R_k^L for every finished product in every region. However, let there be one or more finished products which are considered inessential and upon which for other reasons we do not wish to place availability constraints. For those commodities in any or all regions, we let $R_k^L = 0$. This procedure, in essence, erases the constraints.[68]

A still more general statement of our constraints as given by relations

[66] Note, again, that for all but one J in any region J, $a_{kj}^{J->L} = 0$.

[67] By such a notational change we do not change the meaning of the relation. For when we define R_k^L as a non-positive number, we at the same time multiply it by -1 so that we always use $-R_k^L$ in place of R_k^L when R_k^L was defined as a non-negative number. For example in the previous section if we were to redefine R_4^A as a non-positive number, then in the relations specified in that section we would always replace R_4^A by $-R_4^A$.

[68] It is to be noted that in a linear program **by definition** outputs of finished products can never be less than zero, and exports can never exceed supplies available for export. Hence, the left-hand side of relation (25) can never be greater than zero. Thus when $R_k^L = 0$, the availability constraint on k can never be effectively binding.

(25) is possible. Let the superscript symbol L->L represent L whenever the former occurs below, L = 1,...,U;[69] then relations (25) become:

$$(26) \qquad -\sum_{Jj} a_{kj}^{J->L} x_j^J \leq R_k^L \quad [70] \qquad \begin{matrix} k = 1,\ldots,s \\ L = 1,\ldots,U \end{matrix}$$

The dual to the general linear program defined by relations (24) and (26) may now be stated. For notational convenience, we first let P_k^L replace $\bar{P}_k^L$. Thus P_k^L becomes the relevant subsidy price which is a variable and which is distinct from $\bar{P}_k^L$ which is a constant and is the predetermined finished product price specified as a given for the model. Thus the dual is:

$$(27) \qquad \text{Min} \quad W = \sum_{L=1}^{U} \sum_{k=1}^{s} P_k^L R_k^L$$

subject to

$$(28) \qquad -\sum_{Jk} a_{kj}^{L->J} P_k^J \geq c_j^L \qquad \begin{matrix} j = 1,\ldots,v \\ L = 1,\ldots,U \end{matrix}$$

The objective is, as in the previous section, to minimize fictitious returns to resource owners after deduction of subsidy payments necessary to achieve the required availabilities on finished products.[71] The constraints are also of the same order. Where j is a production activity, all $a_{kj}^{L->J} = 0$ when $J \neq L$, and the relation (28) states that per unit level operation of j the costs of all local inputs less the value of intermediate goods produced less subsidy payments on all finished products produced must not be less than c_j^L, c_j^L being zero when j produces only intermediates. Where j is a dummy activity, all $a_{kj}^{L->J} = 0$ when $J \neq L$, and the relation (28) states that per unit level operation of j the cost of the intermediate or resource having its status shifted less subsidy payment on the finished product must not be less than c_j^L; or, put otherwise, the price of the intermediate or resource less the subsidy price on the finished product must not be less than the predetermined price of the finished product. Finally, when j is a shipping activity all but one $a_{kj}^{L->J} = 0$ when $J \neq L$, and relation (28) may be read in two different ways. When an intermediate good i is shipped, such that $a_{ij}^{L->J} \neq 0$, relation (28) states that per unit level operation of j the costs of all local inputs including the intermediate to be shipped less the value of the output (export) valued at the price prevailing in the region of destination (i.e. region J) must not be less than c_j^L which is zero.

[69] Thus, J->J represents J; U->U represents U; etc.

[70] Recall that when in the summation $J \neq L$, $a_{kj}^{J->L} = 0$ except for that j in J which ships k to L. When in the same summation $J = L$, $X_j^J = X_j^L$; and we have $a_{kj}^{J->L} = a_{kj}^{L->L} = a_{kj}^{L}$ which will be non-zero for some activities in L.

[71] This deduction is effected in the objective function since R_k^L is negative when k runs from (m+1) to s. At this point it should be recognized that the payment of subsidies tends to raise the returns to resource owners above what they otherwise would receive. See Stevens, op. cit.

When a final product good q is shipped, such that $a_{qj}^{L->J} \neq 0$, relation (28) states that per unit level operation of j the cost of all local inputs (excluding the finished product to be shipped) plus subsidy payment on the finished product in region L (the originating region) less subsidy payment on the finished product in region J (the terminating region) must not be less than c_j^L (which equals the spread between the finished product prices in the two regions when the unit level is defined such that $a_{qj}^{L->J} = -a_{qj}^{L->L} = 1$).

At this point, it may be helpful to present the above linear program and its dual in the familiar matrix form. Let

$\mathbf{A} = \left[a_{kj}^{J->L}\right]$, a matrix of technical coefficients of order Us x Uv

$(k = 1,\ldots,s;\ j = 1,\ldots,v$ where $v = n + m + (U-1)r;\ J,L = 1,\ldots,U)$.

$= \begin{bmatrix} A_i \\ A_q \end{bmatrix}$ where A_i is a (Um x Uv) matrix of technical coefficients pertaining to resources and intermediate commodities $(I = 1,\ldots,m)$; and where A_q is a U(s-m) x Uv matrix of technical coefficients pertaining to finished products $(q = m+1,\ldots,s)$

$X = \{X_j^L\}$, a column vector of activity levels of order Uv

$\bar{P}' = \left[\bar{P}_q^L\right]$, a row vector of predetermined finished product prices of order U(s-m)

$c' = \left[c_j^L\right]$, a row vector of per unit level incomes of order Uv

$= \bar{P}' A_q$

$R = \{R_k^L\}$, a Us order column vector of the limiting constants of the constraints pertaining to resource and intermediate commodity use and to availability of finished products; and

$P = \{P_k^L\}$, a Us order column vector of resource and intermediate prices and of subsidy prices or finished products.

The linear program and its dual are:

$$\text{Max } Z = c'X \qquad\qquad \text{Min } W = R'P$$

subject to:

$$-AX \leq R \qquad\qquad -A'P \geq c$$

and

$$X \geq 0 \qquad\qquad P \geq 0$$

11. SOME CONCLUDING, GENERAL REMARKS

A few concluding statements on the general interregional model of the previous section may now be made. First, in selecting a set of levels at which to operate diverse activities, there can be many possible objectives. Maximizing the sum of regional incomes is only one of these. Maximizing

growth rates of one or more regions,[72] maximizing per capita incomes, minimizing transportation costs,[73] minimizing labor inputs,[74] maximizing new employment opportunities, maximizing tax revenue from industrial development are among many others which might be sought. In each case it would be necessary to restate the objective function, define the relevant constraints, identify the most meaningful classification of commodities and list of activities, etc. However, the basic structural elements and procedures of an interregional model would tend to remain the same.[75]

As already indicated, the model has been developed in its purely formal aspects. In practice, an operational model would deviate from the one described in many respects. First, both the number and kind of production activities would vary from region to region. (A coal mining activity need not be listed for a region which completely lacks coal deposits.) Further, the list of shipping activities would vary from region to region. [A coal shipping (export) activity need not be listed for a region which possesses no coal.] Still more, the list of dummy activities would vary from region to region. [The dummy activity converting coal (as intermediate) into coal (as finished product) would not be listed in a tropical region where coal is not used for household purposes.] Additionally, the commodity classification would be different for different regions. (In a static model, iron ore would not appear in the commodity classification of a region which neither possesses deposits of iron ore nor has any steel capacity to consume iron ore.)

Aside from the differences among regions themselves, practical considerations of computation dictate major regional asymmetry. The general interregional linear programming model developed above is non-operational as such. Many too many activities and commodities are envisaged. In any application, the model must be scaled down to a size which can yield numerical results. In scaling down the model, one does not eliminate the same, or even the same number of activities and commodities in each region.

[72] See Moore, **op. cit.**

[73] See Henderson, "A Short-Run Model..." **op. cit.**, and Garrison and Marble, **op. cit.**

[74] See Moses, **op. cit.**

[75] For example, in a transport minimization problem, the objective function would be a summation of the transport inputs of all activities in all regions. The constraints would read "not less than" (that is $\geq$) some limiting constant. Certain constraints would require that deliveries of specific finished products in the several regions be at least as great as a set of prescribed constants. Other constraints would relate to the use of certain limited resources -- such as the fixed capacity of a given rail line, or the fixed industrial plant and equipment of a region, or the fixed capacity of a mine, etc. The later constraints, initially reading "not greater than," would be multiplied through by -1 in order to read "not less than". The particular classification of commodities and set of activities selected would tend, for example, to give less attention to those commodities and activities directly and indirectly requiring little transport inputs such as diamond processing and emphasize those commodities and activities requiring directly and indirectly large amounts of transport inputs, such as steel and cement manufacture, etc. Nonetheless, the basic framework of the model would not be altered.

For certain regions which are off-center and produce specialized products, one may find it desirable to list a relatively small number of activities and commodities, whereas for a central diversified industrial region which tends to use, produce and export many more commodities, one may tend to list a relatively large number of activities and commodities, **ceteris paribus**. Moreover, one tends to utilize other techniques and procedures whereby to cut down an interregional linear program to a manageable size. He may use input-output techniques or comparative cost studies to derive certain magnitudes, or to eliminate certain activities as infeasible, etc. Equally important, an investigator aiming at application must design the model in the light of existing data, possible new sources of data, and resources available for data collection and processing.[76] In this respect, interregional linear programming models confront many of the same problems as do interregional input-output models.[77]

Finally, in this article, no attempt is made to appraise the interregional linear programming technique as such. We have already made explicit some of its shortcomings -- such as the assumption of linear production functions, and the need to specify beforehand finished product prices or the equivalent. In one sense it is too early to judge the promise of this technique. Several points are clear, however. The technique,at least for certain situations, can be of considerable use in establishing competitive norms, and in pointing up inefficiencies in existing patterns of location, transport flows, and resource use.[78] Clear, too, is the fact that the technique is not a completely general one. It is a partial system in a number of important respects. To the extent that it can be developed further -- to embrace household behavior as approximated by consumption functions, agglomeration economies (including economies of scale, localization and urbanization), etc. -- i.e. to the extent that it can approach being a more truly general interregional equilibrium system,[79] and at the same time remain operational, the more effective an optimizing technique it will be. Such progress of course must come via synthesis with the stronger elements of other techniques, such as input-output, industrial complex analysis, and gravity models.

[76]For relevant discussion, see **Report, 1956-57,** Harvard Economic Research Project, Cambridge, Mass., 1957.

[77]In this connection, see W. Isard, "Interregional and Regional Input-Output Analysis: A Model of a Space Economy," **Review of Economics and Statistics,** Vol. 33, Nov. 1951; W. Isard, "Some Empirical Results and Problems of Regional Input-Output Analysis," in W. Leontief, **et. al. Studies in the Structure of the American Economy,** Oxford University Press, 1953; and L. Moses, "The Stability of Interregional Trading Patterns and Input-Output Analysis," **American Economic Review,** Vol. 45, Dec. 1955.

[78]For example, see Henderson, "Efficiency and Pricing...", **op. cit.**

[79]See W. Isard, "General Interregional Equilibrium," **Papers and Proceedings of the Regional Science Association,** Vol.3, 1957; W. Isard and D. Ostroff, "The Existence of a Competitive Interregional Equilibrium," **Ibid,** Vol. 4, 1958; and L. Lefeber, **op. cit.**

AN INTERREGIONAL LINEAR PROGRAMMING MODEL[1]

by Benjamin H. Stevens

1. INTRODUCTION

THE AIMS OF REGIONAL SCIENCE ARE AT LEAST TWOFOLD: (1) AS A SCIENCE, it seeks for the better understanding of regional economic, social, political, and geographic structure and, hence, for better theories of regional development, patterns, and systems; and (2) as a study of regions it seeks better techniques for the empirical analysis of existing patterns and systems.[2] Interregional linear programming falls partly into both of these broad fields of interest. A theoretical linear program can incorporate many of the variables and the interrelationships among these variables

[1]This paper is adapted from the author's doctoral dissertation **Interregional Linear Programming**, which is currently in preparation, and is a revised version of a paper presented at a joint meeting of the Econometric Society and the Regional Science Association in Philadelphia, December, 1957. The author wishes to thank Professor Walter Isard for his invaluable assistance and unflagging inspiration and Professors James M. Henderson and Charles M. Tiebout for their helpful comments. Obviously any shortcomings of this paper are the responsibility of the author alone. The author also wishes to acknowledge support of his research by a grant from Resources for the Future, Inc.

The reader unfamiliar with linear programming may refer to the following items listed in order of increasing difficulty and comprehensiveness: W. Isard, "Interregional Linear-Programming...," this **Journal**, Vol. I, first issue; C.W. Churchman, R.L. Ackoff, and E. Leonard Arnoff, **Introduction to Operations Research**, John Wiley & Sons, New York, 1957, Part V; A. Charnes, W.W. Cooper, and A. Henderson, **An Introduction to Linear Programming**, John Wiley & Sons, New York, 1953; S. Vajda, **The Theory of Games and Linear Programming**, John Wiley & Sons, New York, 1956; R.G.D. Allen, **Mathematical Economics**, St. Martin's Press, New York, 1956, chapters 11-20; R. Dorfman, P.A. Samuelson, and R. Solow, **Linear Programming and Economic Analysis**, McGraw-Hill, New York, 1958; and T.C. Koopmans (ed.), **Activity Analysis of Production and Allocation**, John Wiley & Sons, New York, 1951.

[2]For a discussion of some of the aims and approaches of regional science see W. Isard, **Location and Space-Economy**, John Wiley & Sons, New York, 1956, chapter I and p. 287. See also W. Isard, "Regional Science, the Concept of Region, and Regional Structure," and M.E. Garnsey, "The Dimensions of Regional Science," **Papers and Proceedings of the Regional Science Association**, Vol. II, 1956 and W. Alonso, D. Bramhall, W. Isard, and B. Stevens, "The Core and Boundaries of Regional Science," **Papers and Proceedings of the Regional Science Association, Vol. IV**, 1958.

which determine the economic structure of regions and regional systems; an empirical linear program can handle large amounts of data in a systematic and productive manner.

The general applicability of interregional linear programming, both to the economic theory of regional systems and to the empirical analysis of existing regional patterns, invites comparison of programming models with general spatial equilibrium systems (both continuous and multiregional) at the one extreme and interregional input-output systems at the other. General spatial equilibrium models are highly abstract. The recent "semi-continuous" model of Lefeber[3] and the multiregional model of Isard[4] contain most of the usual conditions of general equilibrium theory plus the added factor of spatial friction. These models extend the one-point general equilibrium systems of Walras,[5] Hicks,[6] and others to include transportation and trade variables. The addition of these variables requires corresponding additions to the equations of the one-point systems. The new equations express conditions on import substitution, on transport input substitution, and on relative prices among spatially separated markets. At the same time, the models maintain the usual competitive equilibrium price conditions governing factor substitution in production and goods substitution in consumption within each spatial market.

At the opposite extreme are the interregional input-output models of Isard and others.[7] These are basically empirical models, identical in form to the non-spatial models of Leontief[8] except that the interindustry matrix is expanded by the inclusion of interregional relationships; each input is specified as coming, not just from a particular industry, but from a particular industry in a specific region. As is usual in input-output, the coefficients in the matrix are assumed constant. Thus fixed proportions of each good are assumed drawn from each region no matter what the pattern of outputs. This, coupled with the usual condition that there be fixed factor proportions in the production of any good (factor proportions in the production of any particular good may vary from region to region but not within a region), makes an interregional input-output system completely deter-

[3]L. Lefeber, **Allocation in Space**, to be published by North-Holland Publishing Company, Amsterdam.

[4]W. Isard, "General Interregional Equilibrium," **Papers and Proceedings of the Regional Science Association**, Vol. III, 1957, pp. 35-60.

[5]L. Walras, **Elements of Pure Economics**, Transl. by W. Jaffe, George Allen and Unwin, London, 1954.

[6]J.R. Hicks, **Value and Capital**, Oxford, Second Edition, 1946.

[7]W. Isard, "Interregional and Regional Input-Output Analysis: A Model of a Space-Economy," **Rev. Econ. Stat.**, Nov. 1951, XXXIII, 318-28. See also, L.N. Moses, "The Stability of Interregional Trading Patterns and Input-Output Analysis," **The American Economic Review**, Vol. XLV, Dec. 1955.

[8]W.W. Leontief, **The Structure of the American Economy**, Oxford University Press, New York, 1951, esp. Part II.

minate. Once a bill of goods is specified, there is no choice mechanism for optimizing shipment patterns or substituting factors to achieve a more efficient allocation of available resources.

Interregional linear programming, as already suggested, falls somewhere between these two extremes. Like general spatial or interregional equilibrium systems it provides a mechanism for choosing optimum shipment patterns and allows continuous substitution between different modes of production of each good (although not between factors in any one productive process). Furthermore this choice mechanism provides a means for determining efficient resource allocations among the various possible productions and shipments.[9] Like interregional input-output it is based on linear homogeneous production functions with constant coefficients. This makes possible the mathematical solution of empirical systems with many industries and many regions. Linear programming shares one real weakness with input-output: the linearity of the production functions precludes any automatic adjustment for economies of scale. Inputs and outputs for each process remain in the same ratios whether the process is employed at unit level or million-unit level. The real effect of allowing only constant-returns-to-scale productive processes is to eliminate the individual firm as the unit of study. In the solution of an interregional linear program, the same production patterns will arise in each region whether many small firms or one giant firm is assumed. We are unable to identify or isolate the activities of the individual firm since there are no scale criteria for establishing firm size.[10]

Perhaps a more damaging weakness of the linear programming technique is that it loses sight of the individual on the consumption-income side as well. There is (as yet) no way to introduce systematically either consumer tastes or curvilinear resource supply schedules into the linear programming framework. The reason for this weakness is implicit in the name "linear"; relationships within the system must be linear with constant coefficients. This means that, were we to introduce households as an endogenous part of the system, they would have to be assumed non-rational, consuming given proportions of goods and supplying given proportions of labor and other resources for all levels of income and prices. Put another way, we would have to assume that households were merely another productive process,

[9] Continuous substitution between different modes of production will not give us continuous factor substitution unless a productive process is specified for every possible combination of factors in the production of each good. Even here we would not have a continuous transformation curve but rather a polygon or segmented curve, each segment expressing factor proportions for a particular process. For a large number of processes, this segmented curve would approximate a continuous curve and for an infinite number of processes could be made exactly equivalent to it; but a linear program with infinite possible processes would be impossible to solve. It is interesting to note, however, that we can approximate continuous transformation curves as closely as we wish, bearing in mind of course that each additional process adds to the difficulty of solving the linear program. In theory, at least, continuous factor substitution is not incompatible with linear programming.

For a more complete discussion of this point, see R. Dorfman et al., op. cit., pp. 348-9.

[10] See R. Dorfman et al., op. cit., Sec. 14-6, pp. 407-8.

supplying outputs of resources in fixed ratio to the final goods they consume.[11]

Such an assumption would add to the complexity of the system without providing us with any commensurate return in increased realism or usefulness of results. We therefore concentrate on the productive sector of the economy, where linear programming is especially applicable, and leave the analysis of the interrelationships in the household sector to techniques which are better suited to the job. Thus the "market", defined as a set of fixed prices for all consumption goods in all regions, is assumed in advance; resources are assumed to be in fixed supply in each region; and any adjustments in consumption, income, and resource supply must be made external to the linear programming framework.[12]

Despite the limited usefulness of linear programming in handling the household sector in a systematic manner, we strongly feel that its ability to analyse the important interrelationships in the productive sector gives it overwhelming advantages for regional analysis. Given somewhat restrictive assumptions about households as consumers of final goods and suppliers of resources, our model is then completely free to vary the mode and level of production in each region, to reallocate resources by making substitutions between different regional supplies, and to adjust shipments toward more efficient patterns.[13] If our aim is the delineation of an optimum

[11] Leontief, (op. cit. p. 42 and pp. 132-6) argues that, empirically, consumers seem to operate in a far more "linear" fashion than might generally be supposed. In particular, he shows that "actual substitution among the different types of consumer expenditures keeps within rather narrow limits." Even he does not claim, however, that there is a fixed relationship between the physical inputs of consumption goods and the physical outputs of labor and other resources in the household sector.

[12] There is still one basic difference between our multiregional model and a typical non-spatial linear program. We do, in fact, consider the individual in the household sector insofar as we specify different "market" prices and resource endowments for our various regions. In this sense, each region is an individual consumer and resource supplier. Thus, we may wish to consider budget constraints, consumption patterns, and factor incomes for our individual regions much as we would if we were handling individual households. Some of the problems arising from these considerations will be discussed in a later section of the paper.

[13] Ways have been suggested for approximating the resource supply and consumption demand relationships in the household sector by a series of iterations external to the linear programming model itself. For example, the size of pool of a consumption good in a region and its assumed market price could be checked against an assumed (or empirical) demand curve for that good. If the supply exceeded the demand, the assumed price could be adjusted downward (and vice versa) and the linear programming problem could be reworked. The test would again be applied and, if necessary, another adjustment would be made, etc. There is as yet no systematic method available for performing this iteration so that it could easily require insupportable amounts of computation. A more complete discussion of this approach can be found in T. Vietorisz, **Regional Programming Models and The Case Study of a Refinery-Petrochemical Synthetic Fiber Industrial Complex for Puerto Rico**, doctoral dissertation, Massachusetts Institute of Technology, January, 1956, Part I.

The author is attempting to attack this problem with a different approach. Results of this attempt should be forthcoming in his dissertation **Interregional Linear Programming.**

multiregional productive economy, linear programming gives us a means toward that end.

The use of linear programming for spatial analysis is not original here. Attempts to generalize the somewhat limited models of von Thunen[14], Weber[15], and later spatial theorists[16] have lead to a variety of spatial linear programming models. For example, the Koopmans-Hitchcock "transportation" model[17] (which can be used for many things besides transportation analysis) was one of the first problems to be solved by linear programming and has become established as a special case of the linear programming technique. Samuelson has employed this "transportation" model in a more generalized form in his pioneering work on spatial price equilibrium.[18] Beckmann and Marschak have applied linear programming to the spatial analysis of the firm and have indicated the extensions necessary to analyse a whole industry.[19] And in a very recent example Koopmans and Beckmann have attacked a plant location problem stressing the difficulty of handling indivisibilities.[20]

Unfortunately none of this work has attacked the central problem for the regional scientist. All of the foregoing models have analysed location with transportation costs assumed fixed; or they have analysed transportation costs with the location of industry assumed fixed. Nowhere until the

[14] J.H. von Thünen, **Der isollerte Staat in Beziehung auf Landwirtschaft und National-ökonomie**, Hamburg, 1826.

[15] A. Weber, **Alfred Weber's Theory of the Location of Industries**, translated by Carl J. Friedrich, Chicago, 1929.

[16] For a history of the development of spatial theory and a description of the work of the better theorists see W. Isard, **Location and Space-Economy**, Chap. II.

[17] F.L. Hitchcock, "The Distribution of a Product from Several Sources to Numerous Localities," **Journal of Mathematics and Physics**, M.I.T., Vol. XX, 1941, pp. 224-30; and T.C. Koopmans, "Optimum Utilization of the Transportation System," **Proceedings of the International Statistical Conference**, Washington, D.C., 1947, Vol. V (reprinted as Supplement to **Econometrica**, Vol. XVII, 1949).

For a simple statement of the transportation problem see S. Vajda, **op. cit.**, p. 31. Koopmans, **Activity Analysis of Production and Allocation** contains chapters on the solution of a transportation problem (Chap. XXIII) and a more complex model of transportation (Chap. XIV).

M. Beckmann, "A Continuous Model of Transportation," **Econometrica**, Vol. XX, 1952, presents a non-linear transportation model which is not applicable within the linear programming framework but does make important contributions to the theory of spatial price equilibrium as related to transport costs.

[18] P.A. Samuelson, "Spatial Price Equilibrium and Linear Programming," **American Economic Review**, Vol. XLII, June, 1952, pp. 283-303.

[19] M. Beckmann and T. Marschak, "An Activity Analysis Approach to Location Theory," **Kyklos**, Fasc. 2, 1955.

[20] T.C. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities," **Econometrica**, Vol. XXV, January, 1957, pp. 53-76.

very recent work of Lefeber have we seen an attempt to construct a linear programming model where both location and transportation costs are allowed to vary, and where transportation is included in the productive system as a consumer of factors[21]. And none of the models has satisfactorily handled the important problem of interregional trade in intermediate goods.

It is our purpose, then, to construct and examine an interregional linear programming model which answers these objections. Certain economic and mathematical features of the model will be discussed, the "dual" or pricing problem will be interpreted, and the possibility of application to empirical regional analysis will be explored.

2. MODEL FRAMEWORK

We will consider a multiregional economy. The concept of region need not concern us here except to note that the type of regional breakdown employed will depend, in empirical models, on the type of data available and the results desired. Bear in mind that regions can, in theory, be as large as a nation or the world in a single-region model or as small as a point in a continuous programming model.

Regions are assumed to be self-contained. That is, transportation within a region is assumed costless, while transportation between regions will ordinarily require transport services. At any given point of time each region is assumed to possess fixed and non-augmentable endowments of natural, capital, and human resources.[1] All resources are considered **immobile** and each must be "extracted" by a suitable productive process which converts it into an intermediate good. These, as well as all other intermediate and final goods, are assumed **mobile.**[2] Thus the **resource coal** in a

[21] L. Lefeber, **op. cit.**

[1] As discussed in section 1, the assumption of strictly fixed resource supplies is necessary in linear programming models. This limitation may not be difficult to accept as applied to natural resources, but may be somewhat disturbing when applied to capital and labor. If we think of the very short run, however, it is not difficult to imagine fixed stocks of capital and labor. A long-run model would require the addition of the dynamic element of capital formation which complicates the model excessively for our purposes. And the addition of thoeries of population distribution and migration would make our model more complete but less workable. For the moment, then, we will confine our attention to the short-run model, keeping in mind the development of more complete dynamic long-run models as a future goal.

[2] The assumption of an "extraction" process for every resource may seem to be an unnecessary complication. It has two overwhelming advantages, however:

(1) It allows us to keep resources (immobile) and intermediates (mobile) strictly separated in the model. This is obviously of less importance in non-spatial models where the mobility of resources is not a vital issue.

(2) It allows us later, in the "dual" problem, to distinguish between the value of a resource "in the ground" and the value of the other resources and intermediate goods required to "extract" it. This will permit us to impute the value of a natural resource back to the land which contains it and to develop some links between linear programming and the theory of rent. See section 4 below.

region must be extracted and converted into the **intermediate product coal** to be used by industry in that region or other regions. A further process is required to convert it into the **consumption good coal**. Similarly, labor and capital are "extracted" by processes which produce labor and capital services as intermediate products. Labor and capital are never consumption goods. However, either or both will enter into the production of final consumption goods such as automobiles, personal services, and leisure.[3] The notation has been specifically designed to keep the resource, intermediate, and final forms of each good separate and distinct.[4]

Intermediate and final goods required in each region are assumed to be drawn from fictitious "pools" of these goods. These pools are formed by shipments into the region from industries in all regions including itself. Thus the pool of steel in the New York region might be made up of shipments from steel plants in Pittsburgh, Cleveland, and Bethlehem as well as local supplies produced by steel plants in Trenton. Note that pools of intermediates and final goods in a region supply only the industries and consumers of that region. Export shipments of goods are made directly from the industries producing those goods, as will be developed below.

Productive processes are posited in the form of linear homogeneous production functions with constant coefficients.[5] These processes may vary from region to region, and within any one region there may be several different processes available for the production of a single good. Such processes can be considered additive; but no single process produces more than one product.[6]

Each process in each region is considered as dissociated into the possible shipments of the product of that process to all regions. Therefore, instead of solving for the outputs of processes we will solve for the shipments of goods from these processes. Thus the solution variables will actually be shipment levels rather than production levels. The total production level for any process in any region can always be found by summing the shipments made from this process to all regions (including itself).

[3]We assume a process in this system whose one input is the intermediate good labor and whose output is the final good leisure. Thus, unless we put a zero market price on leisure, labor will always be used to capacity; those hours of the 24 in each man-day not being used to produce other goods will always be employed in producing leisure.

[4]See section 3 for symbols and notation.

[5]For a discussion of the linear homogeneous production function in linear programming see R.G.D. Allen, **op. cit.**, Chap. XVII, esp. pp. 565-595. See also Koopmans, **Activity Analysis of Production and Allocation, op.cit.**

[6]The assumption of no joint production is not usually a necessary one in linear programming. However, in our particular form of production, with transportation part of the production process and outputs in terms of shipments, it becomes essential to assume that each process has only a single product. See discussion of the shipment model below.

A shipment model of this type obviously requires that transportation be part of the production process since the output of each process is a series of goods deliveries to the various regions. For this reason (and by this method) transport input coefficients are explicitly included in the production matrix. This means that transportation must be supplied from the pool of transport services in the region of origin. However, transport services are an intermediate good and therefore are themselves transportable. The pool of these services in a region can have its source in any or all regions, and no region need produce its own transport inputs.[7]

It is important to note that the dissociation of productions into shipments is equivalent to multiplying the number of processes by the number of possible regions of destination. This means that the input coefficients for a single process must be repeated for each of the possible shipments from this process. The transport input coefficient will be the only one that varies between different shipments from the same process.[8]

At first glance this proliferation of solution variables and their accompanying input coefficients may seem to make the problem inordinately large and complex. If we consider the alternatives, however, we will see that this is not necessarily so. For example, if we adopted the more common linear programming practice of keeping production and shipment variables separate in the model, we would have all the shipment variables above **plus** the variable expressing total production.[9] This would, in effect, give us one **more** variable for each production process. Admittedly it would simplify the constraints expressing the trade, production and consumption balance for each good. But it would also add to the number of these constraints and the difficulty of solving a linear programming problem is even more dependent on the number of constraints than on the number of variables.[10]

[7]The assumption of transportable transport services is a harmless fiction and is necessary if we wish to include transportation in the production process and still have transport services supplied by the region(s) which can produce them most economically.

[8]A typical input coefficient matrix for a region will be found in the Mathematical Appendix. Note for the moment, however, that the actual production function for any good should not vary among different shipments of that good. Steel in Pittsburg, for example, is produced in the same way no matter whether the finished steel is to be shipped to New York, Boston, or any other destination. Only the amount of transportation required to make the shipment will vary. The choice of a particular process for steel **production** is ordinarily not dependent on the destination of the finished steel.

[9]See M. Beckmann and T. Marschak, **op. cit.**

[10]Dantzig, in Koopmans, **op. cit.**, p. 344, points out that for a system with m constraints and n variables, the simplex method requires about $2m^2n$ computations to obtain an optimum solution. It could be argued that, since the dual program contains one constraint for each variable in the direct problem and one variable for each constraint in the direct problem, and since we have the choice of originally solving either the direct problem or the dual, we would always solve, first, the program with the fewer constraints. As will be shown below, the solution of one of the problems (direct or dual) reduces the other to a system of simultaneous linear equations. On this basis, it would not matter whether the direct problem

Another approach which has been used is to specify productions as a series of shipments direct to individual **industries** and **consumers** in each region[11] whereas in our model shipments go to regional **pools** from which local industries and consumers draw. The direct-shipment approach increases the size of the problem only by increasing the number of variables, not the number of constraints. This increase is considerable, however. For example, in a system with **n** possible processes in each region and U regions our model would have a number of variables of order U^2n while the direct-shipment model would have a number of variables of order U^2n^2 (not including shipments to consumers).

It seems clear that shipments are the all-important variables in a spatial model since it is through shipments that we can most easily analyse the possible substitutions between production of a good in one region and its production in another, and between different resource sources and qualities. On the basis of the discussion above we feel that our method expresses these important interrelationships most economically.

We will further diverge from "standard" practice by specifying minimum consumption levels in each region. Generally, in models of entire productive economies, outputs of final goods are left unconstrained. In such models, the aim of the optimization process is to maximize a **vector** of final output quantities or, equivalently, to maximize a scalar sum of values of consumption (GNP). Such an optimization process, unconstrained except for limitations on resource inputs, gives an "efficient" solution where the output of any one consumption good cannot be increased without a commensurate decrease in the output of another.[12] An efficient solution guarantees the "efficient" allocation of resources: i.e. at least one resource will be used up to capacity.[13] Obviously there are a large number of

---

had relatively many constraints or relatively many variables: where it had relatively many constraints, we would merely choose to solve the dual first. Bear in mind, however, that for many applications we may wish to solve **only** the direct problem. Despite the important economic interpretation of the dual, it is the d rect problem which supplies us with the answers which we will most often be seeking in empirical applications of the model: optimum resource allocation, production levels, and shipment patterns. For this reason, we seek to minimize the number of constraints in the direct program.

[11] See L. Lefeber, **op. cit.**

[12] Koopmans, **op. cit.**, pp. 85-86 and esp. theorem 5.6, has proved the equivalence of the vector maximization of final outputs and the scalar maximization of a linear function of final outputs weighted by market prices. See also Dorfman **et al.**, **op. cit.**, section 14-4, pp. 399-404, where a similar theorem is proved. In particular, the latter authors show that "...any time you assign an arbitrary set of prices and solve the linear program of maximizing output you wind up at an efficient point...".

The concept of efficiency, or Pareto-optimality, is explained in both Koopmans, **op. cit.**, p. 60, and Dorfman **et. al., op. cit.**, sec. 14-2, pp. 391-396.

[13] Actually, we will find at least two resources used up to capacity, since labor will be fully utilized at the outset as long as we put a non-zero market price on leisure. See above.

possible efficient solutions for any particular problem, each solution involving capacity limitations on different resources or groups of resources. The particular solution chosen by the linear program will depend on the market prices placed on final consumption goods since these are the weights used to compute the value of GNP.[14]

But simple efficiency may not be the best criterion for our purposes. An efficient solution guarantees efficient resource allocation (with respect to the set of "market" prices assumed) but does not guarantee an optimal distribution of consumption goods among regions. With any particular set of "market" prices, the maximum GNP may be achieved by supplying a surplus of consumption goods to high-market-price regions while leaving other regions without sufficient goods to support even a subsistence standard of living. Once we obtain such an efficient solution we would be free to alter the market price structure and rework the problem to obtain a more satisfactory (or realistic) distribution of final goods. Or in a longrun model, we could assume a redistribution of population with consumers moving from the consumption-poor to the consumption-rich regions.

There are computational economies, however, in specifying the minimum consumption levels necessary to support the assumed labor force for each region because, as we shall see below, the dual programming problem will tell us directly how much prices would have to be raised in the consumption-poor regions to maintain these minimum levels in a free economy. This would preclude the necessity for making a series of market price readjustments and a complementary series of recalculations of the linear program. If, on the other hand, the assumed market prices reflect the actual price structure of a multiregional economy, the dual program will give us the levels of subsidies necessary to bring consumption-poor regions up to subsistence levels.[15]

Finally, we will assume that there is no net production of intermediates in the system as a whole. Since we allow no joint production this is not a difficult assumption; where no process need produce unnecessary intermediates as byproducts, the supply of each intermediate can always be limited to the quantity needed to produce other intermediates and final

[14] The "tracing-out" of the set of efficient points by a series of linear programs is described in Dorfman et al., loc. cit., esp. p. 395.

[15] The presentation and discussion of the dual program will be found in section 4 of this paper.

Note that the specification of minimum consumption levels will force our solution away from an efficient point only insofar as the assumed market price structure does not, of itself, call forth at least minimum quantities of final goods in each region. Where minimum levels would be fulfilled anyway, the minimum consumption constraints become non-binding. Once all minimum consumptions are supplied, the system will continue raising those outputs which add the most to GNP until one or more resource constraints become binding. Thus, to the extent that some goods would have been in sufficient supply in all regions without the minimum consumption constraints, the solution could be said to be efficient. For final goods which would have fallen below minimum consumption levels without the constraints, the solution would not be efficient. Overall, we may wish to call such a solution "semi-efficient".

goods.[16] Furthermore, the system will have no tendency to produce excess intermediates since intermediate products (and resources) are given zero weight in our objective function, GNP.[17] Bear in mind, however, that both intermediates and resources do carry prices (or rents) which are determined within the model and solved for in the dual program.

3. MATHEMATICAL MODEL

Assume **U** regions and **n** goods. Any particular good may exist in resource, intermediate, and final forms. In each form it will carry the same good number but a different identifying letter. Thus r_h, I_h, and f_h ($h = 1,2,\ldots,n$) are the resource, intermediate, and final forms of good **h** respectively.[1] Although we discussed the possibility of multiple productive processes for any particular good in section 2, we will here, for the purposes of simplicity, limit ourselves to a single productive process in each region for each good and will therefore call the processes industries. (The extension to multiple processes can easily be made at the cost of increased complexity of notation). For example, good I_{10} is produced only

[16] The assumption of no joint production is not necessary to the assumption of no net production of intermediates. Recall that we excluded joint production from our model for the purpose of making transportation part of the productive process (see footnote 6 above). Models where joint production is allowed can still carry constraints on net production of intermediates at the cost of some complication. See Koopmans, **op. cit.**, pp. 91-95, and esp. Theorem 5.12.

[17] GNP or gross national product "measures the...(system's)...output of goods and services in terms of its market value." (**National Income**, A Supplement to the Survey of Current Business, United States Department of Commerce, 1951 edition, page 1). On the basis of this definition GNP, as an objective function, has three advantages: (1) it avoids any possibility of double counting by expressing value of output in terms of market prices (market prices on intermediates are zero by assumption); (2) it fulfills the criteria, expressed above, for an efficient or semi-efficient solution; (3) it has high recognition value: the maximization of gross national product should seem a reasonable objective even to readers who are unfamiliar with linear programming or multiregional analysis. For the mathematical formulation of the GNP function see section 3.

[1] Obviously no particular good need actually exist in all three forms. Coal would appear as a resource, intermediate and final good; automobiles as an intermediate and final good, and leisure as a final good only. Note that no good would appear as a resource only since all resources must be converted into intermediates before they can be used in any industry other than the one which "extracts" them (see section 2). Thus iron ore, which is usually thought of as strictly a resource, would be both a resource and an intermediate in our model. Bear in mind that the balancing-out of intermediates will not necessarily mean that all of a resource such as iron ore is used up, but will only mean that there is no net production of the intermediate iron ore. In the solution of the system there may still be some of the resource iron ore left unutilized "in the ground."

The assumption of as many as three or as few as one possible forms for each good will mean that the actual number of goods in the system will be something more than **n**. It will not be necessary to specify this number, however, but only to assume that all goods (and all industries which produce those goods) can exist in all regions. With this assumption of symmetry, the actual number of goods becomes unimportant.

by industry I_{10} (and is the only product of that industry). This industry may exist in all regions, however.[2]

We are now prepared to define, as constants in the model:

$$r_c\, a^L_{i_d} \quad \begin{matrix}(L=1,\ldots,U)\\(c,d=1,\ldots,n)\end{matrix}$$ The number of units of resource **c** required by industry I_d in region **L** to produce a unit output of intermediate **d**.[3]

$$i_d\, a^L_{i_h} \quad \begin{matrix}(L=1,\ldots,U)\\(h,d=1,\ldots,n)\\(d\neq m)\end{matrix}$$ The number of units of intermediate **d** required by industry I_h in region **L** to produce a unit output of intermediate **h**.[4]

$$i_d\, a^L_{f_e} \quad \begin{matrix}(L=1,\ldots,U)\\(d,e=1,\ldots,n)\\(d\neq m)\end{matrix}$$ The number of units of intermediate **d** required by industry f_e in region **L** to produce a unit output of final good **e**.

$$i_m\, a^{LJ}_{i_d}\,,\quad i_m\, a^{LJ}_{f_e} \quad (L,J=1,\ldots,U)\ (d,e=1,\ldots,n)$$ The number of units of transport services required to ship a unit of intermediate **d** and final good **e** respectively from region **L** to region **J**.[5]

[2] In line with the above footnote, there will be as many industries in each region as there are intermediate and final goods. One of the industries in each region will be i_m which is the transportation industry and produces the intermediate transport services, good I_m. There is a separate industry f_m which produces transport services for final consumption, ostensibly by consuming some of I_m.

[3] The specification of units will depend on the good in question. In general, physical goods (coal, steel, automobiles, etc.) will be measured in weight units or numbers; "service" goods (transport services, labor services, etc.) will be measured in combined units such as ton-miles, man-days, and so forth. In all cases the unit of measurement for any particular good will be kept consistent throughout its various forms (resource, intermediate, final) and its various specifications (endowment, production, shipment, consumption).

[4] Specific exclusion here of input coefficients expressing the utilization of transport inputs is due to the special nature of these coefficients. See below and section 2.

[5] Transport input coefficients are expressed in ton-miles per unit shipped. (Shipment units will depend on the good involved). Notice that these coefficients take into account the distance between the region of origin and the region of des-

$E^{L}_{r_c}$ $(L=1,\ldots,U)$ $(c=1,\ldots,n)$ The endowment or maximum locally available supply of resource **c** in region **L**.

$C^{L}_{f_e}$ $(L=1,\ldots,U)$ $(e=1,\ldots,n)$ The assumed minimum level of consumption of final good **e** in region **L**.

$P^{L}_{f_e}$ $(L=1,\ldots,U)$ $(e=1,\ldots,n)$ The given "market" price of final good **e** in region **L**, (expressed in dollars per unit).

Furthermore, we can specify the solution variables for our direct problem:[6]

$L S^{J}_{i_d}$ $(L,J=1,\ldots U)$ $(d=1,\ldots,n)$ The number of units of intermediate **d** produced by industry I_d in region **L** to be shipped from region **L** to region **J**.

$L S^{J}_{f_e}$ $(L,J=1,\ldots,U)$ $(e=1,\ldots,n)$ The number of units of final good **e** produced by industry f_e in region **L** to be shipped from region **L** to region **J**.

Z=GNP The function to be maximized in the "direct" problem.[7]

We will also specify certain "convenience" variables which will aid in the mathematical construction of the model:

---

tination. Thus they can include all the necessary adjustments for different "rates" on different goods and for lower unit "rates" on shipments over greater distances where these apply. The admittedly difficult practical problem of establishing realistic transport input coefficients for the shipment of "service" goods is beyond the scope of this paper.

[6] Dual variables will be specified in the presentation of the dual problem of section 4.

[7] See section 2 for a discussion of the choice of GNP as the objective function.

$X^L_{i_d}$, $X^L_{f_e}$ $(L=1,\ldots,U)$ $(d,e=1,\ldots,n)$ The total levels of output in region L of industries i_d and f_e respectively.

$Q^L_{i_d}$, $Q^L_{f_e}$ $(L=1,\ldots,U)$ $(d,e=1,\ldots,n)$ The sizes of the pools in region L of goods i_d and f_e respectively.

$R^L_{r_c}$, $R^L_{i_d}$ $(L=1,\ldots,U)$ $(c,d=1,\ldots,n)$ The industrial requirements in region L for resource c and intermediate d respectively.

From the foregoing set of constants and variables we can make the following definitions:

$$3.1a \quad \sum_{d=1}^{n} {}_{r_c}a^L_{i_d} X^L_{i_d} = R^L_{r_c} \quad (L=1,\ldots,U) \quad (c=1,\ldots,n)$$

The total requirement by all intermediate industries in region L for resource c (which, since resources are immobile must be supplied from the endowment of region L).

$$3.1b \quad \sum_{h=1}^{n} {}_{i_d}a^L_{i_h} X^L_{i_h} + \sum_{e=1}^{n} {}_{i_d}a^L_{f_e} X^L_{f_e} = R^L_{i_d} \quad (L=1,\ldots,U) \quad (d=1,\ldots,n) \quad (d\neq m)$$

The total requirement by all industries in region L for intermediate d (supplied from the pool of d in L).

$$3.1c \quad \sum_{J=1}^{U} \sum_{h=1}^{n} {}_{i_m}{}^L a^J_{i_h} {}^L s^J_{i_h} + \sum_{J=1}^{U} \sum_{e=1}^{n} {}_{i_m}{}^L a^J_{f_e} {}^L s^J_{f_e} = R^L_{i_m} (L=1,\ldots,U)$$

The total requirement by all industries in region L for intermediate m, transport services (supplied from the pool of m in L).

Where

$$3.2a \quad X^L_{i_d} = \sum_{J=1}^{U} {}^L s^J_{i_d} \quad (L=1,\ldots,U) \quad (d=1,\ldots,n)$$

Total output of i_d in L is defined as the total of all shipments of intermediate i_d from industry i_d to the pools of all regions including L.

$$3.2b \quad X^L_{f_e} = \sum_{J=1}^{U} {}^{L}s^{J}_{f_e} \quad \begin{matrix}(L=1,\ldots,U)\\(e=1,\ldots,n)\end{matrix}$$ Total output of f_e in L

is defined as the total of all shipments of final good f_e from industry f_e to the pools of all regions including L.

And

$$3.3a \quad Q^L_{i_d} = \sum_{J=1}^{U} {}^{J}s^{L}_{i_d} \quad \begin{matrix}(L=1,\ldots,U)\\(d=1,\ldots,n)\end{matrix}$$ The size of the pool of

I_d in L is defined as the sum of all shipments of I_d into region L from all regions including L.

$$3.3b \quad Q^L_{f_e} = \sum_{J=1}^{U} {}^{J}s^{L}_{f_e} \quad \begin{matrix}(L=1,\ldots,U)\\(e=1,\ldots,n)\end{matrix}$$ The size of the pool of

f_e in L is defined as the sum of all shipments of f_e into region L from all regions including L.

Our constraint system[8] can now be written in the simple form:

$$3.4a \quad R^L_{r_c} \leq E^L_{r_c} \quad \begin{matrix}(L=1,\ldots,U)\\(c=1,\ldots,n)\end{matrix}$$ The utilization of resource

r_c by industries in L cannot exceed the endowment of c in L.

$$3.4b \quad R^L_{i_d} = Q^L_{i_d} \quad \begin{matrix}(L=1,\ldots,U)\\(d=1,\ldots,n)\end{matrix}$$ The pool of intermediate

I_d in L will be used up by industrial consumption in L.

$$3.4c \quad C^L_{f_e} \leq Q^L_{f_e} \quad \begin{matrix}(L=1,\ldots,U)\\(e=1,\ldots,n)\end{matrix}$$ The pool of f_e in L must

be at least large enough to supply the minimum consumption of f_e in L.

Rewriting constraints 3.4 in standard linear programming form with constraining constants on the right[9] we can form:

[8]For discussion of systems of linear constraints, see any of the references on linear programming cited in section 1. See section 2 for the considerations underlying the formulation of constraints 3.4 and their forms 3.5, 3.7, etc.

[9]This is merely a convention and is not essential to the logic of the model. However, this particular form helps to simplify the computational problem.

$$3.5a \quad R^L_{r_c} \leq E^L_{r_c}$$

$$3.5b \quad R^L_{i_d} - Q^L_{i_d} = 0 \qquad (L=1,\ldots,U) \quad (c,d,e=1,\ldots,n)$$

$$3.5c \quad -Q^L_{f_e} \leq -C^L_{f_e}$$

which are the basic constraints of our model in its simplest form.

$$3.6 \quad Z = \sum_{L=1}^{U} \sum_{e=1}^{n} P^L_{f_e} Q^L_{f_e} + \sum_{L=1}^{U} \sum_{e=1}^{n} (0)\, Q^L_{i_d} \qquad (L=1,\ldots,U)$$ [10]

then our "direct" problem is to maximize the function 3.6 subject to the constraints 3.5, and to non-negative values for any pool and/or shipment variable.[11]

Taking this simple form of our problem as a basis and substituting from 3.1 to 3.3, we are able to form:

$$3.7a \quad \sum_{d=1}^{n} r_c a^L_{i_d} X^L_{i_d} \leq E^L_{r_c} \qquad (L=1,\ldots,U) \quad (c=1,\ldots,n)$$

$$3.7b(1) \quad \sum_{h=1}^{n} {}_{!_d} a^L_{!_h} X^L_{i_h} + \sum_{e=1}^{n} {}_{i_d} a^L_{f_e} X^L_{f_e} - \sum_{J=1}^{U} {}_J S^L_{i_d} = 0 \qquad (L=1,\ldots,U) \quad (d=1,\ldots,n) \quad (d \neq m)$$

[10] The measurement of GNP by total value of consumption pools may seem strange at first. Recall, however, that all final goods actually produced (all net production in the system) are shipped into the consumption pools in the various regions for distribution to consumers in those regions. Since pool sizes are measured before consumption (and since all final goods in pools are assumed to be consumed), total value of consumption pools is a perfectly acceptable measure of GNP.

Note that the pools of intermediates appear in the GNP function. Since they are weighted by their zero market prices and have no affect on the value of Z, they could just as well be left out. However, they are included for the sake of symmetry and because they appear in the complete matrix presentation of the model (see Mathematical Appendix).

[11] The requirement that all variables be non-negative is a simple mathematical necessity in linear programming. This would have to be true in our model in any case since we specified that shipments are from the industry in the region of origin. Such shipments would obviously not be reversible and the concept of imports as negative exports would be both unnecessary and mathematically impossible in our model.

$$3.7b(2)\quad \sum_{J=1}^{U}\sum_{h=1}^{n} {}_{i_m}a^{J}_{i_h}\,{}^{L}s^{J}_{i_h} + \sum_{J=1}^{U}\sum_{e=1}^{n} {}_{i_m}a^{J}_{f_e}\,{}^{L}s^{J}_{f_e} - \sum_{J=1}^{U} {}^{J}s^{L}_{i_m} = 0 \qquad (L=1,\ldots,U)$$

$$3.7c\quad -\sum_{J=1}^{U} {}^{J}s^{L}_{f_e} \leq -c^{L}_{f_e} \qquad \begin{matrix}(L=1,\ldots,U)\\(e=1,\ldots,n)\end{matrix}$$

Finally, by substituting from 3.2 and rewriting we can form the constraints:

$$3.8a\quad \sum_{J=1}^{U}\sum_{d=1}^{n} {}_{r_c}a^{L}_{i_d}\,{}^{L}s^{J}_{i_d} \leq E^{L}_{r_c} \qquad \begin{matrix}(L=1,\ldots,U)\\(c=1,\ldots,n)\end{matrix}$$

$$3.8b(1)\quad \sum_{J=1}^{U}\sum_{h=1}^{n} {}_{i_d}a^{L}_{i_h}\,{}^{L}s^{J}_{i_h} + \sum_{J=1}^{U}\sum_{e=1}^{n} {}_{i_d}a^{L}_{f_e}\,{}^{L}s^{J}_{f_e} - \sum_{J=1}^{U} {}^{J}s^{L}_{i_d} = 0 \qquad \begin{matrix}(L=1,\ldots,U)\\(d=1,\ldots,n)\\(d\neq m)\end{matrix}$$

$$3.8b(2)\quad \sum_{J=1}^{U}\sum_{h=1}^{n} {}_{i_m}a^{J}_{i_h}\,{}^{L}s^{J}_{i_h} + \sum_{J=1}^{U}\sum_{e=1}^{n} {}_{i_m}a^{J}_{f_e}\,{}^{L}s^{J}_{f_e} - \sum_{J=1}^{U} {}^{J}s^{L}_{i_m} = 0 \qquad (L=1,\ldots,U)$$

$$3.8c\quad -\sum_{J=1}^{U} {}^{J}s^{L}_{f_e} \leq -c^{L}_{f_e} \qquad \begin{matrix}(L=1,\ldots,U)\\(e=1,\ldots,n)\end{matrix}$$

and the objective function:

$$3.9\quad Z = \sum_{J=1}^{U}\sum_{e=1}^{n} {}^{J}s^{L}_{f_e}\,p^{L}_{f_e} + \sum_{J=1}^{U}\sum_{d=1}^{n} {}^{J}s^{L}_{i_d}\,(0)$$

$$^{J}s^{L}_{i_d},\ {}^{J}s^{L}_{f_e} \geq 0,$$

which provide the complete statement of our "direct" problem.[12]

Some explanation of constraints 3.8 is in order. Constraint 3.8a states that the utilization of a resource in local production of intermediates in a region for shipment to that region and to other regions must not exceed the local endowment of that resource. This is the expected relationship since resources are assumed strictly non-transportable and since an intermediate phase is inserted between the resource and final forms of a good so that consumption goods are not produced directly from resources.

The first two terms of 3.8b (1 and 2) give the industrial demands for intermediates to produce and transport all shipments to other regions. The third term gives the available supply of intermediates in the local pools. Our assumption of exact balance of supply and demand for intermediates follows from implicit efficiency considerations in the model which preclude the production of excess amounts of intermediates.

Finally, 3.8c expresses the requirement that the total amounts of each consumption good shipped into the pool of a region must at least meet the minimum consumption level for that good in that region. This requirement raises a question of consistency. Our model is bounded on both ends. At one end we are constrained by limited resource endowments; at the other end, by the need to provide minimum consumptions. It is at least theoretically possible that these two sets of constraints (3.8a,c) could be inconsistent; there might not be sufficient resources in any particular system to produce even the minimum required quantities of all consumption goods. Were this the case it would be indicated by a failure to find the required non-negative values for the variables (S's) at an optimal solution. The alternatives would be to reduce the minimum consumption levels and rework the problem, or eliminate them entirely and work for a strictly efficient solution. For theoretical purposes, however, this is no real objection since we can always assume that there are sufficient resources among the regions as a whole to produce the final (and necessary intermediate) goods and transport them to market in such a way as to fulfill minimum consumption levels.

If we accept the foregoing assumptions, the optimization problem presents no great difficulties. A technique such as the simplex method[13] would be used to find the maximum value of the objective function (3.9) and the corresponding values for the S's. As is typical in linear programming, we would find in an empirical problem that one or more resources would be used up to their limits. The constraints for these particular resources

[12] Note that we have given no count of constraints. Such counts are unimportant in linear programming since the equivalence of numbers of variables and constraints is in no way a necessary condition for solubility. Actually, the number of variables will greatly exceed the number of constraints in most models. For example, 3.8 constraints $3Un$ constraints in $2Un^2$ variables. See section 3, footnote 1.

[13] Koopmans, op. cit., Chap. XXI, XXII, XXIII, gives a very complete description of the mathematical theory and practice of this computation method. Simpler descriptions are to be found in Vajda, op. cit., and Charnes, Henderson, and Cooper, op. cit.

would then be "binding."[14] Other resources would be available in excess of needs. Similarly, certain of the minimum consumptions in some regions might just barely be supplied and their constraints, too, would be binding. Associated with these binding constraints will be scarcity rents which we shall discuss in section 4.

It may still not be entirely clear how resources are allocated within the model.[15] If we take, for example, the final good automobiles (say good f_g) in region V we can think of the system as first checking the minimum consumption level specified by $C^V_{f_g}$. (We will assume, for the moment, that all other minimum consumptions are already being met). From the left-hand side of 3.8c we (and the system) observe that f_g would have to be produced in region V or in one or more other regions for shipment to region V. The system, in theory, would then substitute possible shipments of automobiles ($^Ls^V_{f_g}$ for all L) in constraints 3.8b(1) to determine how much of I_g and other intermediates would be required in each region to produce these shipments. The system would also check through constraints 3.8b(2) to see how much of transport services (and by extension, how much of the intermediates required to produce transport services including, perhaps, automobiles themselves) would be needed to make these possible shipments. The required primary intermediates (including transport services) in each region would have to be produced locally or imported from other regions as indicated by the third terms of 3.8b(1) and (2).

The system would then check on the secondary intermediates required to produce the primary intermediates as shown in the first terms of 3.8b(1) and (2) and on the tertiary intermediates required to make these secondary intermediates, etc. The possible local productions and imports of all these intermediates in each region ultimately require the use of resources as prescribed in constraints 3.8a. If certain regional intermediate productions required resources in excess of endowments, the possible productions and shipments of these intermediates to supply industry f_q would have to be eliminated from consideration. Since the necessary resources are available somewhere,[16] however, the intermediates and hence the final outputs of automobiles could be produced (bearing in mind that the more devious the shipment route from resource to final good, the more intermediates that would be used up in transportation).

Automobiles would then be shipped until the left-hand side of constraint 3.8c were equal to the right. At the same time, Z (eq. 3.9) would be increased by an amount equal to the total of these shipments multiplied by $P^V_{f_g}$. Recall that the manner in which the model supplies f_g, the intermediates which its production and shipment requires, and the resources

[14] For a discussion of the concept of binding constraints, see Charnes, Henderson, and Cooper, op. cit.

[15] The following discussion is completely hypothetical since production and shipment levels are actually determined simultaneously within the model. However, such a discussion may serve to elucidate in a **ceteris paribus** manner some of the choices which are actually made in the model **mutatis mutandis**.

[16] Recall our two assumptions: (1) The system is consistent; sufficient resources exist to provide at least all minimum consumptions. (2) All minimum consumptions, other than automobiles in region V, are already being met.

needed to produce those intermediates does not affect the value of the minimum consumption level of automobiles supplied to region V. Thus any combination of productions and shipments could be used to furnish this supply so long as resource endowments were not exceeded.

If, in supplying all minimum consumptions in all regions, no resource is used up to capacity, we know that we do not yet have an optimum solution; and even if some resource constraint is binding,[17] it may be possible to produce more of certain outputs without increasing the demand on the scarce resource.[18] In either of these two situations we are in a position to increase the value of GNP. In theory, at least, the system would now try all possible reallocations of productions and rearrangements of shipment patterns in the attempt to make GNP as large as possible, still subject to the requirement that all minimum consumption levels be met. Obviously the number of possible choices would be fantastically large, but the simplex method of solution provides the model with a systematic way of making the best choices. From both mathematical proof and past experience with linear programming models (simpler ones, of course, but with exactly the same principles and objectives involved) we know that the outcome is assured even if we find it difficult actually to describe the myriad interrelated choices which are considered in reaching that outcome.[19]

Our result will be in the form of a maximum value for Z (GNP) and non-negative values for all the S's. Many, if not most, of the S's will have a zero value. This is not surprising when we consider that there will probably be a best region for the supply of each non-local good, at least within each subgroup of regions. Because of the differences in regional production functions and resource endowments it is likely that one region will be able to produce a particular non-local good and ship it to nearby regions more economically (in terms of resource utilization) than the receiving regions can produce it.[20] Thus, for any region which does not actually produce a certain good, S's expressing shipments of that good from that region will all be zero. As explained above, relationships 3.2

[17] Keep in mind that the constraint on labor is always binding (see section 2). This means that any decision to use more labor in production requires a reduction of the amount of labor used in leisure. If we attach to leisure a non-zero market price, the system will need to test the value of GNP to determine whether the market value of any new production will at least equal the value lost in reducing the output of leisure.

[18] For example, the further production of men's suits may be possible in some region even though the constraint on a highly specialized resource such as titanium is binding in all regions.

[19] The outcome is generally assured even if degerenacy is encountered in the simplex computations. See Dorfman et al., op. cit., pp. 92-93, for a discussion of the degeneracy problem and methods for circumventing it in applications of the simplex method.

[20] This would not be the result of scale or agglomeration economies in spatially concentrated production since such economies are not compatible with linear homogeneous production functions and hence are not considerations in our model. See section 1, above.

can be used to determine total production levels. Hence, the optimum pattern for production and trade within the whole economy can be determined.

Further implications of the model are discussed in section 5. There we also note the type of results expected in certain applied forms of the model.

4. THE DUAL PROBLEM[1]

It has been demonstrated that every direct linear programming problem has a dual problem associated with it. But the dual is more than a mathematical formality; rather, it is a representation of a set of relationships contained in the model which, like the relationships in the direct problem, have, at least in part, a real economic interpretation. There is an expression common to linear programming which states that if the direct program is an allocation problem, the dual is a valuation problem. In particular, if we are allocating resources among the production and shipment activities of the various regions as in the program 3.8-3.9, our dual is the problem of imputing the market value of final goods back to the ultimate resources used in producing those goods.

The objective of the dual is to minimize the total of these imputed values; at the minimum point, we could call this total NI (national income) and it will be equal (as we might expect) to the maximum value of GNP found in our direct problem above.[2] But there is some difficulty in providing a realistic interpretation for this total at levels above the minimum since, at these higher levels, it would represent a national income not actually realized by the owners of resources.[3]

[1] The formal mathematics of converting the direct problem into its dual are much easier to show in matrix algebra form. We will therefore dispense with many of the formalities here, and present the detailed manipulations in the Mathematical Appendix. A discussion of the mathematical and theoretical relationships between the direct and dual problems can be found in any of the references cited at the beginning of Section 1. See esp. Vajda, **op. cit.**, Chapter VIII.

[2] "Total output is also measured, in terms of the factor costs of producing it, by the national income--the aggregate earnings of labor and property which arise from...production." **National Income, loc. cit.** On the basis of this definition, and by the argument of Section 2 above, NI seems a reasonable interpretation of the dual objective function. The equivalence of GNP as a measure of value of output and NI as a measure of value of inputs is basic to national income accounting in the general sense.

It is further to be noted that the equivalence of our optimal values of GNP and NI is guaranteed by the fact that a direct program and its dual have a common optimal solution or no solution at all. Thus we have the choice of formally solving (by simplex or other method) either the direct or dual problem although, in any case, we will usually wish to determine the variable values for both problems. See below.

[3] The logical difficulty of interpreting the dual objective function as NI at any level above the optimum is discussed in the previous article by Isard in this journal. Suffice it to say, for the moment, that our NI function is defined in the same mathematical sense as Isard's **Fictitious Returns to Resource Owners** and at an optimal solution is identical to his **Effective Returns to Resource Owners.**

The above difficulty, in itself, is interesting. The selection for the dual variables of values which are too high (and which thus make our NI total higher than its optimum value) will mean that factor costs exceed revenues from consumption for most industries. Then production by and shipment from these industries will not take place, and the resources (either pure or in the form of intermediates) needed to support these industries will not, in fact, be utilized.

Hence, the dual of the linear program attacks the problem of placing values on resources (and, as a corollary, on intermediates and the supplying of minimum consumptions) in such a way as to guarantee the efficient (or semi-efficient) utilization of those resources. Since efficient (or semi-efficient) resource allocation is the nominal aim of the direct problem, the dual is then a pricing system, internal to the linear program, which sets prices such that the results of the direct problem would be achieved even if all decisions were made independently by individual firms seeking to maximize profits on the basis of market prices and factor costs.[4]

This result is not surprising if we recall that the linear programming technique requires the assumption of a purely competitive economy.[5] Thus we should expect that "excess" profits would not exist in the system. With these considerations in mind, we will find the formulation of the dual constraints quite logical. If our dual variables, which we call "shadow" prices or rents,[6] are defined as:

$u^L_{r_c}$ $(L=1,\ldots,U)$ $(c=1,\ldots,n)$ — The rent imputed to resource c in region L;

$v^L_{i_d}$ $(L=1,\ldots,U)$ $(d=1,\ldots,n)$ — The shadow price of intermediate d in region L;

$w^L_{f_e}$ $(L=1,\ldots,U)$ $(e=1,\ldots,n)$ — The location utility rent imputed

[4]For a discussion of attainment of efficiency under a regime of decentralized decisions, see Koopmans, op. cit., pp. 93-96. Koopmans shows that within the linear programming framework, the same efficient allocation will be achieved whether decisions on activity levels are made by a central planning authority at one extreme or a myriad of individual producers at the other.

[5]See Dorfman, et al., op. cit., pp. 407-408.

[6]Both terms, "shadow" price and rent, are used interchangeably. We prefer, however, to think of rents as pertaining to scarcities (roughly their usual economic interpretation) and of shadow prices as pertaining to internal accounting prices of the system. We therefore define the dual variables for intermediates as shadow prices and the dual variables for resources and scarce final goods as rents.

to final good e in region L;[7]

we can then interrelate these variables in constraints expressing the fact that "excess" profits cannot exist. Or, put otherwise, no activity (production, shipment, etc.) can earn positive profits.[8] For all activities which are **not** carried on (i.e. have zero values for the solution variables in the direct problem) profits will be zero or negative. For all activities which are carried on, profits will be exactly zero.[9] It is a further rule that for every variable in the direct problem, we have a constraint in the dual (and for every constraint in the direct problem, we have a variable in the dual).[10] Therefore we must write a non-positive profitability constraint for every possible shipment. These constraints will have the form:

$$4.1a \qquad \sum_{c=1}^{n} r_c a^L_{i_d} u^L_{r_c} + \sum_{\substack{h=1 \\ h\neq m}}^{n} i_h a^L_{i_d} v^L_{i_h} + {}^{L}_{i_m}a^{J}_{i_d} v^L_{i_m} \geq v^J_{i_d} \qquad \begin{matrix}(L,J=1,\ldots,U) \\ (d=1,\ldots,n)\end{matrix}$$

The sum of the values of all the resources and intermediates used to produce a unit output of I_d in region L, plus the value of the transportation

[7]Although rents on consumption goods are unusual in linear programming models, they appear here because we have specified minimum consumptions and constraints requiring that these minimum consumptions be supplied in the direct problem. See below for the interpretation of location utility rents and their use in determining subsidy levels for consumption-poor regions.

[8]"Normal" profits can exist in the sense that the amount paid for capital and entrepreneurial or managerial ability can be considered profits. These returns to capital and organizational skills appear only because inputs of such capital and skills are specified in the technology of the direct problem above. "Excess" profits would then be the difference between revenue and cost where cost included all such items as fair return to investment and managerial ability.

[9]Therefore, once one problem, either the direct or dual, is solved by some systematic technique such as the simplex method, the values of the variables in the other are easily determined. This is because for each binding constraint (i.e. where the "=" holds) in the solution to one problem, the corresponding variable usually takes a non-zero value in the other. Similarly, for each non-binding constraint (i.e. where the "<" or ">" holds) in the solution to the one problem, the corresponding variable always takes a zero value in the other. Once we have the solution to one of the problems, then, we know which variables may take non-zero values and which constraints are binding in the other (and from the previous discussion, we also know the optimal value of the objective function). We can then eliminate non-binding constraints and zero variables to obtain a straightforward, if lengthy, set of simultaneous linear equations to solve. A more complete discussion of these points can be found in Dorfman et al., op. cit., Chapter 3.

[10]Vajda, loc. cit. Note that dualism is a completely reversible process. The dual of the dual of a direct problem is the direct problem itself.

required to ship a unit of I_d to region J, must be greater than or equal to the shadow price of I_d in region J.

$$4.1b \quad \sum_{\substack{d=1 \\ d\neq m}}^{n} {}_{i_d}a^{L}_{f_e} v^{L}_{i_d} + {}_{i_m}a^{L}_{f_e} v^{L}_{i_m} \geq w^{J}_{f_e} + P^{J}_{f_e} \qquad \begin{matrix}(L,J=1,\ldots,U) \\ (e=1,\ldots,n)\end{matrix}$$

The sum of the values of all the intermediates used to produce a unit output of f_e in L, plus the value of the transportation required to ship that unit of f_e to J, must be greater than or equal to the sum of the location utility rent and the market price of f_e in J.

In the simplest sense, constraints 4.1 state that the production cost of a unit of a good in one region plus the cost of transporting that unit to another region, must be at least as great as the price of a unit of that good in the region of destination. This is the expected relationship for spatial price equilibrium. We know from the work of Samuelson and other economic theorists (and from intuition) that if the price spread between two regions **exceeds** the transport costs, it will pay a producer in the lower-price region to continue shipping goods to the higher-price region until supply and demand conditions in the two regions make the price spread exactly **equal** to the transport costs. Then further shipments are unprofitable. Note, on the other hand, that wherever the price spread is less than or equal to the transport costs at the outset, no shipments will be made at all.[11]

Now recall that transport costs **within** a region are zero:

$$\begin{matrix} {}_{i_m}a^{L}_{i_d} \\ , \\ {}_{i_m}a^{L}_{f_e} \end{matrix} = 0 \qquad \begin{matrix}(L=1,\ldots,U) \\ (d,e=1,\ldots,n)\end{matrix}$$

Then constraints 4.1 express the non-profitability conditions on intraregional as well as interregional shipments. In particular, they state that the shadow price of an intermediate in a region must be no greater than the value of the inputs used to make it, and that location utility rent plus the market price of a final good in a region must sum to no more than the value of the inputs used to make that good.

Now to complete our mathematical presentation of the dual, we rewrite our constraints 4.1, once more putting the constraining constants on the right.

$$4.2a \quad \sum_{c=1}^{n} {}_{r_c}a^{L}_{i_d} u^{L}_{r_c} + \sum_{\substack{h=1 \\ h\neq m}}^{n} {}_{i_h}a^{L}_{i_d} v^{L}_{i_h}$$

[11] Samuelson, op. cit.

$$+ \; {}_{i}{}^{L}_{m}a^{J}_{i}{}_{d} \; v^{L}_{i}{}_{m} - v^{J}_{i}{}_{d} \geq 0 \qquad \begin{matrix}(L,J=1,\ldots,U)\\(d=1,\ldots,n)\end{matrix}$$

$$4.2b \quad \sum_{\substack{d=1\\d\neq m}}^{n} {}_{i}{}^{L}_{d}a^{L}_{f}{}_{e} \; v^{L}_{i}{}_{d} + {}_{i}{}^{L}_{m}a^{J}_{f}{}_{e} \; v^{L}_{i}{}_{m} - w^{J}_{f}{}_{e} \geq P^{J}_{f}{}_{e} \qquad \begin{matrix}(L,J=1,\ldots,U)\\(d=1,\ldots,n)\end{matrix}$$

Then we can form our objective function, $NI=Z'$, as:

$$4.3 \quad Z' = \sum_{c=1}^{n}\sum_{L=1}^{U} E^{L}_{r}{}_{c} \; u^{L}_{r}{}_{c} + \sum_{d=1}^{n}\sum_{L=1}^{U} (0) \; v^{L}_{i}{}_{d} + \sum_{e=1}^{n}\sum_{L=1}^{U} (-C^{L}_{f}{}_{e}) \; w^{L}_{f}{}_{e}$$

$$u^{L}_{r}{}_{c}, \; v^{L}_{i}{}_{d}, \; w^{L}_{f}{}_{e} \geq 0$$

Note that the rents and shadow prices in the national income function (4.3) are weighted by the corresponding constants from the constraints of the direct problem (3.8).[12] Thus the shadow prices on intermediates are given zero weight which is what we should expect. Any value imputed to an intermediate is automatically reimputed back to the resources used to make it (as can be traced in 4.1a). Therefore, a positive weighting of intermediate accounting prices would be, in effect, double counting. Bear in mind, however, that intermediate shadow prices do have an important interpretation. The shadow price of an intermediate in a region will be the "equilibrium" price on the pool of that intermediate in that region. That is, it will be the price that users of the intermediate in the region must pay for it, and the price which producers of the intermediate (in any or all regions) must meet if they hope to ship into the pool.

There are several differences between our interpretation of resource rents and that ordinarily found in non-spatial linear-programming models. First, recall that we specify an extraction process for **every** resource. Therefore, the "cost" of converting a resource into its primary mobile intermediate includes the value of all the intermediates utilized in the extraction process (and through them, the value of the resources used to make the intermediates). For this reason, we can interpret our **u**'s as pure rents in the sense that they apply only to the resource "in the ground" and **not to its extracted value.** Thus the "cost" of coal to a steel producer includes as two separate items: (1) rent on the resource coal, and (2) the "cost" of extracting the resource and converting it into the intermediate coal.

[12]Also part of the rules of the linear programming game. See Vajda, loc. cit.

Second, we think of resource rent itself as composed of two separate items: royalty and rent. Following Marshall, we define royalty as the amount "when accurately adjusted, (which) represents the diminution of the value of the mine, regarded as a source of wealth in the future."[13] Because ours is a short run model, and because the dual program provides non-zero rent values only for resources used up to capacity, we will find all royalty values to be zero. In essence, there is no provision in our model for depletion of stock. We point to the concept of royalty, however, because it would appear in a dynamic model where it would be important to separate it from rent. But in our static short-run model rent as such reduces to a scarcity concept.

Third, we choose to separate our resource scarcity rents into two parts by degree of scarcity: absolute rent on scarcity from the point of view of the system as a whole, and differential rent expressing differences in scarcity among regions. We think of a resource as absolutely scarce when it is used up to capacity in all regions and relatively scarce when it is used up to capacity only in one or a few regions. Then we define the lowest regional rent value for a particular resource as its absolute rent; for resources which are not scarce from the point of view of the system as a whole, (i.e. when in at least one region, the resource endowment exceeds industrial requirements) this value will be zero; for resources which are scarce in every region, this value will generally be greater than zero. Additionally, we define the difference between the absolute rent on a resource and its actual rent (u_r) in a particular region as the differential rent on the resource in that region. Thus the total rent $u^L_{r_c}$ of resource c in region L is the sum of its absolute rent, $\left[\text{min. } u^J_{r_c} \quad (J=1,\ldots,U)\right]$ and its differential rent, $\left[u^L_{r_c} - \text{min. } u^J_{r_c}\right]$.

Under these definitions, absolute rent is a single value for each resource. But differential rent will usually vary from region to region for any particular resource. The problem remains of making the most useful interpretation of the differential rents which we extract from the solution to our dual.

In the classical theory, differential rent arises because different qualities of land are used to produce a certain product, say wheat, which carries a single price in the competitive market. Better quality land will be able to produce as much wheat per unit with less labor and capital (or more wheat per unit with the same amount of capital and labor) as land of inferior quality. Then the lowest quality land actually in use for wheat production will earn zero rent (zero return over and above the cost of capital and labor) while better quality lands will receive differential rents depending upon their differences in quality. Zero will be the lowest level to which rent can fall since owners will withdraw land from production rather than earn negative rents.[14]

[13] A. Marshall, **Principles of Economics**, Seventh Edition, Macmillan, London, 1916, p. 439.

[14] Thorough presentations of the classical theory of rent can be found in D. Ricardo, **The Principles of Political Economy and Taxation**, Dutton, New York, 1911; J.S. Mill, **Principles of Political Economy**, Longmans Green, London, 1909; and Marshall, **op. cit.**

If we extend this concept to resources other than land, we can see that the lowest quality resource deposit actually "mined" will earn zero differential rent, while better quality deposits (easier to "mine") will earn non-zero differential rents. And if, as several authors have suggested, we think of "quality" as composed of **both** fertility in the general sense and location, we can interpret **our** differential rents within the classical framework as arising from differences in quality.[15]

But suppose we go one step further and assume that for any particular resource r_c, the production function for the primary intermediate I_c to be extracted from that resource is identical in every region. This is the equivalent of saying that in every region all deposits of r_c are of the same fertility, that the same proportions of capital, labor, and other factors are used for the production of one unit of I_c from r_c. Then if differential rents still appear, they can only be interpreted as location rents.[16]

This is certainly not surprising. If two equally fertile deposits of r_c are both in use for the supply of a given market, the one which is closer to the market should earn a higher return since less transportation will be required to deliver its primary intermediate, I_c, to the market. The problem of interpreting such location differential rents remains, however, because in our linear program we are dealing with many markets, one in each region, and not with a single market in the classical sense. Hence, the two deposits of r_c in the example above may be supplying entirely separate markets. Then differential rents on the two deposits cannot be directly interpreted as measures of locational advantage in the supply of any **particular** market.

This difficulty is by no means insurmountable. Only lack of space prevents a thorough development of the strong ties between existing location theory and our interregional linear programming via location rents. From the early work of von Thunen and Weber up to the more recent studies by Losch, Hoover, Isard, and Dunn, location rent has been a basic element in the theory of both agricultural and industrial location; and the relationship between transport costs, location choices, and location rents has been both recognized and analyzed.[17] Then, if our direct linear program is an appropriate generalization of certain partial location models, it should not be surprising that we can extract from the dual many of the same price and rent relationships found in these models.[18]

[15]Location as an element of quality in the determination of rents is stressed in Weber, **op. cit.**, Editor's Introduction, and is mentioned in Ricardo, **op. cit.**, p. 221; Mill, **op. cit.**, p. 433; and Marshall, **op. cit.**, p. 441.

[16]See esp. Mill, **loc. cit.** and Marshall, **loc. cit.** for the introduction of the concept of location or "situation" rent.

[17]Appropriate references to these authors can be found above and in Isard, **Location and Space-Economy, op. cit.** Especially noteworthy is E.M. Hoover, Jr., **Location Theory and the Shoe and Leather Industries**, Harvard, Cambridge, 1937, Chapter II.

[18]A much more complete discussion of this point will be found in the author's dissertation as referred to above.

Finally, we are left with the problem of interpreting the w's, or location utility rents, which are imputed to final goods. Ordinarily in linear programming rents on final goods do not appear. Recall, however, that we purposely applied minimum consumption constraints to final goods in our direct problem. Then by our rule that we must have one variable in the dual for each constraint in the original problem, we perforce must place some sort of imputed value on final goods. But more important, we should expect some value to be attached to the supplying of minimum consumption levels if, in a purely efficient solution, these consumptions would not be supplied.

We find that this is actually the case; where minimum consumption constraints are binding we generally find non-zero values for the w's in the dual. For the direct problem, we can interpret these binding constraints to mean that the system is being forced away from an efficient towards a semi-efficient solution in order to supply the binding consumptions. For the dual problem, it is probably better to think of the w's as subsidies paid either directly to consumers or to the producers of certain final goods in order to elicit minimum supplies of these goods in consumption-poor regions. This interpretation is almost automatic if we look at constraints 4.1b where the w's appear as additions to the market prices for the final goods. They have the effect of raising the given market prices in consumption-poor regions to a level where those regions can command minimum consumption supplies.

Since the value of the subsidies is imputed back to the owners of ultimate resources (as can be traced through constraints 4.1 or 4.2), we could, if we wished, think of the system as being governed by a welfare administrator (perhaps an interregional or federal government) who would tax the resource owners just enough to provide sufficient revenues to make the subsidy payments. It would be a mistake to think, however, that this tax on rents to provide subsidies would necessarily involve no dead-weight loss to the system. Actually, it can be shown that although the value of GNP=NI may remain the same, more usually it falls when minimum consumption constraints are applied to the maximum GNP problem (and hence, location utility rents are added to the minimum NI problem). Therefore, the decision to apply minimum consumption constraints (and provide subsidy payments) involves welfare considerations of both efficiency and distribution.[19]

[19]Since the value of the subsidies is imputed back to the resource owners, and since the u's appear with positive coefficients and the w's with negative coefficients in the objective function, 4.3, it is tempting to think of them as balancing out. Certainly the value of the subsidies is balanced by their imputed value to resource owners. But if we start with an efficient solution, with consumptions unconstrained, the addition of minimum consumption constraints will usually require the reallocation of resources. And reallocation, itself, will have its own affect on the value of the u's in addition to the effects caused by the imputation of subsidies. The net of these effects will ordinarily be to reduce the value of the objective function. But this is not surprising. If we recall the discussion of efficiency in section 2, we should expect the value of GNP=NI to be lower in the semi-efficient than in the efficient solution.

This point may still not be entirely clear, however. Although it is obvious that resources which are not scarce in an efficient model may become scarce when minimum consumptions are forced, the converse is not nearly so obvious. But suppose that in an efficient solution two resources are scarce (for at least some regions

Finally, the appearance of location utility rents in the dual has a further value. Recall that one of the weaknesses of the linear program is its inability to handle systematically the household sector of the economy. Approaches have been suggested (and are being explored) for making market price adjustments after each solution of the program so that, upon successive recalculation and readjustment, supply and demand for each final good in each region will tend to equalize. Now if we assume for the moment that the market prices and the minimum consumption levels are originally set with reference to realistic demand schedules, the location utility rents will give us an indication of how our market prices should be adjusted to provide the desired consumption levels.[20] Most simply, if we add the value of each **w** to its corresponding **P** and recompute the program, all consumptions should then be fulfilled with all new **w**'s zero.[21]

Obviously, such an extensive readjustment would not ordinarily provide an "equilibrium" solution. Market prices would now tend to be too high, for the consumption levels provided, in the originally consumption-poor regions. Probably an upward adjustment of market prices in consumption-poor regions by only a portion of their **w**'s together with a corresponding downward adjustment of the prices in consumption-rich regions would lead most

---

their **u**'s ≥ 0). The reallocation of one of them, say, to producing minimum consumptions in a semi-efficient solution would reduce the production of some other goods (or the production of the same goods for some other regions). If these other goods shipments would have required large amounts of the second resource, the total endowments of this resource might no longer be utilized in a semi-efficient solution (and the value of its **u** in some of the regions which were supplying it might fall to zero). Since the "extra" rent on the first resource will tend to be balanced by the subsidy payments, the net effect will be to reduce GNP=NI.

There are two possible cases in which the addition of minimum consumption constraints will still allow an efficient solution with the value of GNP=NI unchanged: (1) where minimum consumptions are already being supplied without the minimum consumption constraints so that none of these constraints are binding and all **w**=0; and (2) where minimum consumptions are not being supplied but there exist multiple optimal solutions of the efficient program, at least one of which fulfills all minimum consumptions. Here one or more of the minimum consumption constraints would probably be binding, but all **w** would still be equal to zero.

Since we are talking about digressions from an efficient solution, it is easy to show that the addition of minimum consumption constraints could not possibly raise the value of GNP=NI.

[20] A discussion of this problem was introduced in Section I of this paper. Note that here we think of minimum consumption levels, not as subsistence levels, but as minimum acceptable levels at a given set of market prices. Thus a consumer research analysis would tell us what quantity of each good would be bought (other goods prices remaining constant) at each market price. If we pick a price for each good and use its corresponding demand quantity as our minimum consumption level, we can achieve the situation outlined here. The problem of the interrelatedness of demand schedules must, perforce, be sidestepped in this rough approximation.

[21] It should be noted that if the **w** and **P** on the right-hand side of constraints 4.1b were added to form a new $\mathbf{P}^1$, the **w** having already been found for a first solution, the new solution would be efficient in the new set of market prices even if the original solution were only semi-efficient.

rapidly to a consumer equilibrium.[22] In any case, the location utility rents provide a first approximation to accurate market price adjustments and may help reduce the number of program recalculations necessary to achieve an equilibrium.

This approach is of little help on the resource-supply side of the household sector, however. Adjustments of the fixed resource endowments to fit computed rent patterns within a framework of completely inelastic resource supply schedules would be a thankless task at best. A more rational approach would be to take regional income levels into consideration when picking price-quantity points on the various final good demand schedules. If this approach does not seem very rigorous, it should be remembered that a linear program is not a general equilibrium system. Adjustments which must be made external to the model depend exclusively on the ingenuity and judgment of the analyst.[23]

5. CONCLUSIONS

The interregional linear programming model which we have constructed has been presented only in its general form. This was our intention from the start. But it would be well to present a brief glimpse of some practical applications of the model.

Students of the linear programming technique who are also interested in spatial problems will note that this model (3.8, 3.9, and 4.2, 4.3) contains within it many of the existing spatial programming models. With suitable assumptions we can reduce our model to solve the Koopmans-Hitchcock problem, the Beckmann and Marschak problem, and many others. But we do not create a general spatial model to solve specific problems. We construct it for its value in the understanding of the spatial economic system, and for its applicability to problems which the more specific models cannot solve.

For example, we could apply the model to a multi-regional analysis of the U.S. economy if we made certain assumptions as to regional resource endowments, minimum consumptions, etc. Then a solution to the problem would give us a picture of an optimum space-economy which we could compare with the existing location and shipment pattern.

This comparison would have major value since it would point up inefficiencies in the present pattern. With the assumption that the economy

[22] The question of optimal price adjustments is still very definitely open to study and discussion. The author is currently developing a method of simulating demand schedules with step-type functions within the model in the hope of increasing the speed and accuracy of the adjustment process. Since such a method requires substantial additions to the constraint system, the problem remains of balancing the time lost in program calculation with the time saved in the adjustment process.

[23] Unfortunately, there does not appear to be any way to introduce into the linear program itself the requirement that regional incomes equal (or exceed) regional expenditures. Such a requirement would have to be in the form of constraints which would involve both final goods shipment variables (and their prices) and resource rent variables (and their endowment quantities). There is no way of constructing a workable linear program with variables from both direct and dual problems appearing in the constraints of one of those problems.

is reasonably competitive, we would expect these inefficiencies to gradually disappear through relocation of economic activity and/or of population. The comparison would give us the likely direction and size of movement and provide the planners of the various regions with estimates of growth or decline due to regional shifts. Or, where inefficiencies are strongly established, it would give the planners a firmer basis for influencing (for example, through planning controls) future locational decisions.

Furthermore, an adaptation of this model could be used to analyse the differential regional effects of a major locational decision such as the building of the Fairless Steel Works. Or it could be used to study the effects of the new interstate highway system on the existing pattern of production and interregional trade. And bear in mind that the model can be used, within limits, to study breakdowns of regions within regions even to the level of the metropolitan area.

It goes without saying, however, that the model which is only sketched above is merely a mechanism. It must be provided with data and, more important, with a welfare criterion. The goal of the optimization process, whether it is maximization of GNP, minimization of social cost, or whatever, must be established by the analyst. This raises the admittedly difficult questions of data collection (and processing) and goal-setting. But, in theory at least, the possible applications of this type of model are limited only by the ingenuity of the researcher and the digestive power of the computing machine.

MATHEMATICAL APPENDIX

The formal mathematics of linear programming are readily available in the literature and ordinarily would require no elaboration in a short paper of this type.[1] However, the somewhat unusual nature of the interregional linear program, both in the size and construction of the coefficient matrix and in the number and arrangement of the variables, demands a more complete formal presentation than has already been given. The reduction of both the direct and dual problems to matrix math form will demonstrate that our interregional program is merely a more elaborate form of the standard linear program, subject to the same mathematical rules and susceptible to the same computation techniques as any other linear program. It will also show the close relationship between the direct and dual problems, a relationship which tends to suffer loss in emphasis in a linear program as complex as ours.

Throughout this presentation all variables, coefficients, and the like retain the definitions given them in sections 3 and 4 above. In general, matrices and vectors will be ascribed identifying letters which relate closely to their contents so that the thread of the argument may be easily followed. We begin by prescribing the following coefficient matrices:

[1] Refer to any of the works cited in footnote 1, section 1. See esp. Koopmans, **op. cit.**, Chapter III.

$$
{}_{i}^{L}A_{i}^{J} = \begin{bmatrix}
{}_{i_1}a_{i_1}^{L} & {}_{i_1}a_{i_2}^{L} & \cdots & {}_{i_1}a_{i_n}^{L} \\
{}_{i_2}a_{i_1}^{L} & {}_{i_2}a_{i_2}^{L} & \cdots & {}_{i_2}a_{i_n}^{L} \\
\vdots & \vdots & & \vdots \\
{}_{i_m}a_{i_1}^{J} & {}_{i_m}a_{i_2}^{J} & \cdots & {}_{i_m}a_{i_n}^{J} \\
\vdots & \vdots & & \vdots \\
{}_{i_n}a_{i_1}^{L} & {}_{i_n}a_{i_2}^{L} & \cdots & {}_{i_n}a_{i_n}^{L}
\end{bmatrix} \qquad (L,J=1\ldots U)
$$

Dimension: n x n
Number of matrices of this type: U^2

$$
{}_{i}^{L}A_{f}^{J} = \begin{bmatrix}
{}_{i_1}a_{f_1}^{L} & {}_{i_1}a_{f_2}^{L} & \cdots & {}_{i_1}a_{f_n}^{L} \\
{}_{i_2}a_{f_1}^{L} & {}_{i_2}a_{f_2}^{L} & \cdots & {}_{i_2}a_{f_n}^{L} \\
\vdots & \vdots & & \vdots \\
{}_{i_m}a_{f_1}^{J} & {}_{i_m}a_{f_2}^{J} & \cdots & {}_{i_m}a_{f_n}^{J} \\
\vdots & \vdots & & \vdots \\
{}_{i_n}a_{f_1}^{L} & {}_{i_n}a_{f_2}^{L} & \cdots & {}_{i_n}a_{f_n}^{L}
\end{bmatrix} \qquad (L,J=1\ldots U)
$$

Dimension: n x n
Number of matrices of this type: U^2

$$ {}_{r}A_{i}^{L} = \begin{bmatrix} r_1 a_{i_1}^{L} , r_1 a_{i_2}^{L} , \cdots , r_1 a_{i_n}^{L} \\ r_2 a_{i_1}^{L} , r_2 a_{i_2}^{L} , \cdots , r_2 a_{i_n}^{L} \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ r_n a_{i_1}^{L} , r_n a_{i_2}^{L} , \cdots , r_n a_{i_n}^{L} \end{bmatrix} \qquad (L=1,\ldots U) $$

Dimension: n x n
Number of matrices of this type: U

If we now specify $\underline{0}$ as the null matrix of order n, we can form the following matrices:

$$ {}_{i}\bar{A}_{i} = \begin{bmatrix} {}_{1}^{1}A_{i}^{1}, \underline{0}, {}_{1}^{1}A_{i}^{2}, \underline{0}, \cdots, {}_{1}^{1}A_{i}^{U}, \underline{0}; \cdots; \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0} \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0}; \cdots; {}_{i}^{U}A_{i}^{1}, \underline{0}, {}_{i}^{U}A_{i}^{2}, \underline{0}, \cdots, {}_{1}^{U}A_{i}^{U}, \underline{0} \end{bmatrix} $$

$$ {}_{i}\bar{A}_{f} = \begin{bmatrix} \underline{0}, {}_{i}^{1}A_{f}^{1}, \underline{0}, {}_{1}^{1}A_{f}^{2}, \cdots, \underline{0}, {}_{1}^{1}A_{f}^{U}; \cdots; \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0} \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0}; \cdots; \underline{0}, {}_{1}^{U}A_{f}^{1}, \underline{0}, {}_{i}^{U}A_{f}^{2}, \cdots, \underline{0}, {}_{i}^{U}A_{f}^{U} \end{bmatrix} $$

$$ {}_{r}\bar{A}_{i} = \begin{bmatrix} {}_{r}A_{i}^{1}, \underline{0}, {}_{r}A_{i}^{1}, \underline{0}, \cdots, {}_{r}A_{i}^{1}, \underline{0}; \cdots; \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0} \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \cdot \cdot \cdot \cdot \quad \cdot \; \cdot; \cdots; \cdot \cdot \cdot \cdot \quad \cdot \cdot \\ \underline{0}, \underline{0}, \underline{0}, \underline{0}, \cdots, \underline{0}, \underline{0}; \cdots; {}_{r}A_{i}^{U}, \underline{0}, {}_{r}A_{i}^{U}, \underline{0}, \cdots, {}_{r}A_{i}^{U}, \underline{0} \end{bmatrix} $$

Dimension: (of each of the above matrices) $Un \times 2U^2n$
Number of matrices of each of the above three types: 1

If $\mathbf{I}$ is defined as the identity matrix of order Un, and $\bar{0}$ is the null matrix of order Un, we are prepared to define:

$$B_i = \left[\mathbf{I}, \bar{0}, \mathbf{I}, \bar{0}, \ldots, \mathbf{I}, \bar{0}\right]$$

$$B_f = \left[\bar{0}, \mathbf{I}, \bar{0}, \mathbf{I}, \ldots, \bar{0}, \mathbf{I}\right]$$

each of dimension $Un \times 2U^2n$. Then to form our direct constraints, we need only specify the following column vectors:

$$E = \begin{Bmatrix} E^1_{r_1} \\ E^1_{r_2} \\ \vdots \\ E^1_{r_n} \\ E^2_{r_1} \\ \vdots \\ E^2_{r_n} \\ \vdots \\ E^U_{r_n} \end{Bmatrix}$$

Dimension: $Un \times 1$

$$Q_i = \begin{Bmatrix} Q^1_{i_1} \\ Q^1_{i_2} \\ \vdots \\ Q^1_{i_n} \\ Q^2_{i_1} \\ \vdots \\ Q^2_{i_n} \\ \vdots \\ Q^U_{i_n} \end{Bmatrix}$$

Dimension: $Un \times 1$

$$C = \begin{Bmatrix} C_{f_1}^{1} \\ C_{f_2}^{1} \\ \vdots \\ C_{f_n}^{1} \\ C_{f_1}^{2} \\ \vdots \\ C_{f_2}^{2} \\ \vdots \\ C_{f_n}^{U} \end{Bmatrix} \qquad \text{Dimension: } Un \times 1$$

$$Q_f = \begin{Bmatrix} Q_{f_1}^{1} \\ Q_{f_2}^{1} \\ \vdots \\ Q_{f_n}^{1} \\ Q_{f_1}^{2} \\ \vdots \\ Q_{f_n}^{2} \\ \vdots \\ Q_{f_n}^{U} \end{Bmatrix} \qquad \text{Dimension: } Un \times 1$$

$$R_r = \begin{Bmatrix} R_{r_1}^{1} \\ R_{r_2}^{1} \\ \vdots \\ R_{r_n}^{1} \\ R_{r_1}^{2} \\ \vdots \\ R_{r_n}^{2} \\ \vdots \\ R_{r_n}^{U} \end{Bmatrix} \qquad \text{Dimension: } Un \times 1$$

$$R_i = \begin{Bmatrix} R_{i_1}^{1} \\ R_{i_2}^{1} \\ \vdots \\ R_{i_n}^{1} \\ R_{i_1}^{2} \\ \vdots \\ R_{i_n}^{2} \\ \vdots \\ R_{i_n}^{U} \end{Bmatrix} \qquad \text{Dimension: } Un \times 1$$

$$S = \begin{Bmatrix} {}^{1}S_{i_1}^{1} \\ {}^{1}S_{i_2}^{1} \\ \vdots \\ {}^{1}S_{i_n}^{1} \\ {}^{1}S_{f_1}^{1} \\ \vdots \\ {}^{1}S_{f_n}^{1} \\ {}^{1}S_{i_1}^{2} \\ \vdots \\ {}^{1}S_{f_n}^{2} \\ \vdots \\ {}^{1}S_{i_1}^{U} \\ \vdots \\ {}^{1}S_{f_n}^{U} \\ {}^{2}S_{i_1}^{1} \\ \vdots \\ {}^{2}S_{f_n}^{1} \\ {}^{2}S_{i_1}^{2} \\ \vdots \\ {}^{2}S_{f_n}^{U} \\ \vdots \\ {}^{U}S_{i_1}^{1} \\ \vdots \\ {}^{U}S_{f_n}^{U} \end{Bmatrix} \qquad \text{Dimension: } 2U^2n \times 1$$

We are now prepared to write:

X.1a $_{r}\bar{A}_{i}\ S = R_{r}$ which is equivalent to the left-hand side of 3.8a and to 3.1a;

X.1b $_{i}\bar{A}_{i}\ S + {}_{i}\bar{A}_{f}\ S = R_{i}$ which is equivalent to the first two terms of 3.8b and to 3.1b;

X.2a $B_{i}\ S = Q_{i}$ which is equivalent to 3.3a;

X.2b $B_{f}\ S = Q_{f}$ which is equivalent to 3.3b.

Since, from 3.5, we can write:

$$R_{r} \leq E$$

$$R_{i} - Q_{i} = \emptyset^{2}$$

$$-Q_{f} \leq -C$$

we can then substitute and form:

X.3a $_{r}\bar{A}_{I}\ S \leq E$

X.3b $_{i}\bar{A}_{i} + {}_{i}\bar{A}_{f} - B_{i}\ S = \emptyset$

X.3c $-B_{f}\ S \leq -C$

which are equivalent to constraints 3.8. By a process of adjoining we can rewrite X.3 as:

X.4a
$$\begin{bmatrix} {}_{r}\bar{A}_{i} \\ {}_{i}\bar{A}_{i} + {}_{i}\bar{A}_{f} - B_{i} \\ -B_{f} \end{bmatrix} \left\{ S \right\} \leq \left\{ \begin{matrix} E \\ \emptyset \\ C \end{matrix} \right\}$$

or, more simply,

X.4b $D\ S \leq G$ where D is the matrix of dimension $3Un \times 2U^{2}n$ which premultiplies S and G is the vector of dimension $3Un \times 1$ to the right of the inequality sign in X.4a.[3]

[2] $\emptyset$ is the null vector of dimension $Un \times 1$.

[3] Although previously and in X.3b we required that intermediates exactly balance out, multiplying out X.4a would give us the result $_{i}\bar{A}_{i} + {}_{i}\bar{A}_{f} - B_{i}\ S \leq \emptyset$ which implies that intermediates need not balance. No real difficulty is encountered, however. In the simplex computation procedure, when inequalities are made into equalities through the addition of slack variables, no such variables would be added to the intermediate product constraints. They would be treated as equalities just the way they stand.

Now in order to present the objective function of the direct problem, we need only specify:

$$P = \begin{Bmatrix} P^1_{f_1} \\ P^1_{f_2} \\ \vdots \\ P^1_{f_n} \\ P^2_{f_1} \\ \vdots \\ P^2_{f_n} \\ \cdot \\ \cdot \\ P^U_{f_n} \end{Bmatrix}$$

Dimension: $Un \times 1$

Then, from 3.6,

X.5a $\beta' Q_i + P' Q_f = Z$. By proper adjoining of vectors we can rewrite X.5a as:

X.5b $\left[\beta', P'\right] \begin{Bmatrix} Q_i \\ Q_f \end{Bmatrix} = Z$ and by substituting from X.2, we have:

X.5c $\left[\beta', P'\right] \begin{pmatrix} B_i & S \\ B_f & S \end{pmatrix} = Z$ or calling $\begin{Bmatrix} B_i \\ B_f \end{Bmatrix}$ simply B, we have:

X.5d $\left[\beta', P'\right] B S = Z$. If we now define $F' = \left[\beta', P'\right] B$, then

X.6 $F' S = Z$, where F' is a vector of dimension $1 \times 2U^2n$ and Z is a scalar. X.6 is the equivalent of 3.9. Our direct problem, then, is to maximize X.6 subject to X.4b and all $S \geq 0$.

The construction of the dual problem is now very easy once we specify H, a vector of rents and shadow prices (see next page)[4]. Then we know that our direct problem, X.4b and X.6, must have the dual

X.7a $G' H = Z'$ to be maximized subject to

X.7b $D' H \geq F$ and all $u, v, w \geq 0$. This is the equivalent of 4.3 and 4.2,

[4]The appropriate schematic relationsnip between the direct and dual can be found in Koopmans, op. cit., Chapter XIX.

and of course, $Z' = Z$ since the transpose of a scalar is the same scalar. Performing the transpose of D we can rewrite X.7b as:

$$\text{X.8a} \quad \left[({}_r\bar{A}_i)', ({}_i\bar{A}_i + {}_i\bar{A}_f - B_i)', (-B_f)'\right]\{H\} \geq F$$

If we performed the matrix multiplication in the definition of F' above and then found F, and if we substituted for H, we could rewrite X.8a as:

$$\text{X.8b} \quad \left[({}_r\bar{A}_i)', ({}_i\bar{A}_i + {}_i\bar{A}_f - B_i)', (-B_f)'\right]\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} \geq \begin{Bmatrix} \beta \\ P \\ \beta \\ P \\ \vdots \\ \beta \\ P \end{Bmatrix}$$

Multiplying out and disjoining gives:

$$\text{X.9a} \quad ({}_r\bar{A}_i)' u + ({}_i\bar{A}_i - B_i)' v \geq \beta_n$$

$$\text{X.9b} \quad ({}_i\bar{A}_f)' v - (B_f)' w \geq P_n$$

where β_n and P_n are column vectors of dimension U^2n formed by adjoining "U" β's and "U" P's respectively. X.9 a,b is then the close equivalent of 4.2 a,b.[5]

$$u = \begin{Bmatrix} u_{r_1}^1 \\ u_{r_2}^1 \\ \vdots \\ u_{r_n}^1 \\ u_{r_1}^2 \\ \vdots \\ u_{r_n}^2 \\ \vdots \\ u_{r_n}^U \end{Bmatrix} \qquad v = \begin{Bmatrix} v_{i_1}^1 \\ v_{i_2}^1 \\ \vdots \\ v_{i_n}^1 \\ v_{i_1}^2 \\ \vdots \\ v_{i_n}^2 \\ \vdots \\ v_{i_n}^U \end{Bmatrix} \qquad w = \begin{Bmatrix} w_{f_1}^1 \\ w_{f_2}^1 \\ \vdots \\ w_{f_n}^1 \\ w_{f_1}^2 \\ \vdots \\ w_{f_n}^2 \\ \vdots \\ w_{f_n}^U \end{Bmatrix}$$

Dimension: Un x 1 Dimension: Un x 1 Dimension : Un x 1

[5]Carefel study of the relevant matrices will convince the reader that in rows of D' where ${}_r\bar{A}_i'$, ${}_i\bar{A}_i'$, and B_i' contain non-zero elements, ${}_i\bar{A}_f'$ and B_f' contain only zero elements and vice versa. Therefore the result of the matrix multiplication and disjoining shown in X.9 is perfectly legitimate.

$$H = \begin{Bmatrix} u \\ v \\ w \end{Bmatrix}$$

Dimension: 3Un x 1

Similarly, X.7a becomes:

X.10a $\left[E^t, \beta^t, -C^t\right] \begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = Z^t$. Multiplying this out gives:

X.10b $E^t u + \beta^t v - C^t w = Z^t$ which is the close equivalent of 4.3.

Unfortunately, lack of space prevents us from presenting these matrix manipulations in anything more than schematic form. It is hoped, however, that this appendix will enable the interested reader to follow the thread of the program construction.[6] It will have to be accepted on faith that detailed multiplication of the appropriate matrices and vectors will reproduce **exactly** the individual constraints of the direct and dual problems of sections 3 and 4.

The direct (X.4b and X.6) and dual (X.7a and X.7b) problems comprise the standard form of most linear programs. Therefore, they can be solved by the simplex or other linear programming computational technique. This point requires neither proof nor further discussion here.[7]

[6] The similarities (and differences) between our program form and that of Koopmans, **loc. cit.**, p. 328-9, should be noted. Reference to Koopmans original formulation of the linear programming problem in the presence of intermediate commodities, **op. cit.**, Chapter III, is also encouraged.

[7] Ibid, Chapters XXI and XXII.

PHYSICS OF POPULATION DISTRIBUTION

by John Q. Stewart and William Warntz

This report describes on-the-average regularities which have been proved to exist in the distribution of people within cities and in rural areas and across countries as a whole. A review is included of a good deal of previously published work, but new data are presented as well. In particular, none of the tables and figures has appeared in print before, although many of the tables are assembled from social physics notebooks of eight or ten years ago -- when this sort of approach had not achieved its current degree of acceptability.

The "social mechanics" principles indicated in the present paper are necessarily inadequate to deal with other wide areas of human behavior which also have significance for social science. But then mechanics is only one aspect of physics. Space here is not available to describe how, by developing ideas about "social energies," the approach becomes powerfully broadened.

1. Population Densities within a City

Every city tends to conform to a common pattern of internal population distribution, although a variety of local disturbing factors may in individual cities obscure or severely modify it. Not a single city exists which would be so free from these as to exhibit exactly the standard pattern. The standard city is circular in shape. The density of population, **D**, within it is maximum at the center where it is D_o. The density drops radially from the peak in all directions, according to the exponential formula:

$$D = D_o \, 2^{-r/b},$$

where **r** is the distance from the center, and **b** is a short distance which is a constant for the given city.

We call **b** the "halving distance," because every increase in **r** by that amount carries to a ring where the density is cut in half. Thus contours of equal density are concentric circles. Their common center is the center of the "residential city." (In the next section we shall consider the "occupational city".)

TABLE 1. Evidence for the Relation of the Area to the Population of U.S. Cities, 1940.

Rank of city	Log C	Rank of city	Log C	Rank of city	Log C
1	2.67	31-35	2.44	151-165	2.61
2	2.59	36-40	2.48	166-180	2.60
3	2.62	41-45	2.41	181-195	2.55
4	2.47	46-50	2.45	196-210	2.65
5	1.98	51-55	2.54	211-225	2.54
6	2.60	56-60	2.41	226-240	2.56
7	2.55	61-65	2.58	241-255	2.56
8	2.64	66-70	2.54	256-270	2.61
9	2.76	71-75	2.59	271-285	2.48
10	2.65	76-80	2.51	286-300	2.57
11	2.58	81-85	2.63	301-315	2.52
12	2.70	86-90	2.64	316-330	2.53
13	2.70	91-95	2.48	331-345	2.63
14	2.73	96-100	2.62	346-360	2.48
15	1.98	101-105	2.69	361-375	2.52
16-20	2.49	106-120	2.58	376-390	2.51
21-25	2.37	121-135	2.55	391-405	2.46
26-30	2.52	136-150	2.68	406-412	2.55

Log **C** is the logarithm to base 10 of **C**, where $C = P^{3/4}/A$, **P** being the population of any city and **A** the land area in square miles within its political limits (Census of 1940). The rank in the first column is the order of their population size, New York being rank 1, Chicago 2, etc. Los Angeles, rank 5, is notorious for the excessive area within its city limits. The largest 15 cities are listed individually. After that only medians are listed -- of groups of 5 cities each to rank 105, then of groups of 15 for the smaller cities (84323 to 25087). Where medians are tabulated, **P** is the median population of the group and **A** the median land area: as a rule these values do not then refer to the same city.

The same rule, the proportionality of area to the three-fourths power of population, held for the 140 metropolitan districts (listed in the 1940 census for all cities above 50,000); but for these the value of **C** in the formula is only 45 instead of 357.

In addition, five more values of log **C** were computed for medians of five groups of, respectively, 11, 11, 11, 41, 41 small cities each. These have median populations of 11087, 8186, 6253, 4134, 2577 and median land areas of 3.9, 2.3, 1.8, 1.3, 1.2 sq. mi. For these groups, log **C** was found to be, respectively, 2.44, 2.57, 2.59, 2.60, 2.48. This shows satisfactory agreement with the formula for cities as small as 2500. With the inclusion of these five, 59 values of log **C** are listed. Their average is 2.553, corresponding to $C = 357$, and $A = P^{3/4}/357$. (Data compiled by Catherine Kennelly.)

In 1940 and earlier the standard city had a definite edge at some radius, **a**, where the edge density, **D'**, was

$$D' = D_o\, 2^{-a/b},$$

D' being constant for all cities in the United States, and equal to about

TABLE 2. Peak Densities for Tracted Cities, 1940.

Number of Cities	Median Population	Median Peak Density
1	7,455,000	264,000
3	1,931,000	71,000
5	859,000	90,000
7	587,000	48,000
11	368,000	32,000
13	282,000	28,000
13	152,000	35,000
7	86,000	29,000

From city maps published by the Census for the 60 cities for which population by census tracts was listed in 1940, the tract of highest population density in each city was determined from estimation of tract areas. These densities are presented in the form of medians for small groups of cities arranged by size of city. The 60 individual cities ranged in size from that of New York to that of Macon, Ga., (population 58,000). The median densities of the densest tracts are given in people per square mile of the tract area (areas of streets not deducted). Scales of miles published with the census maps of cities were in some instances suspect -- indeed the computed density for the peak tract of Atlantic City, N.J., was at first so out of line with the general relation that the published scale was checked on the ground and found to be in error. The resultant correction eliminated the discrepancy.

Other scattered inspections on the ground in the 1940's indicated that the peak densities for typical smaller cities likewise tended to decrease regularly with population, to about 4000 per square mile for cities of 2500, the census lower limit. Thus U. S. Cities in 1940 tended to conform to a common pattern of internal population distribution, with peak density (in units of persons per square mile) roughly equal to 75 times the square root of a city's total population. But see also Table 4.

2000 persons per square mile. The city's total population, P_c, is the integral of the density, D, taken over the whole area out to the boundary radius, a, successive ring by ring from center to edge. The area, A, of course is πa^2

The first indication of this pattern of concentric rings decreasing exponentially in population density followed from the observation that on the average the area, A, of a city was proportional to the three-fourths power of the population, P_c.[1] Table 1 presents the original statistics. The only reasonable mathematical formula for internal density distribution which can yield such a relation between area and a power of the population is the exponential one for D as a function of r which is given above.

Then study of 1940 census data for all tracted United States cities, as well as inspection of several actual cities, indicated the value of D'. To fit the three-fourths power rule the halving distance, b, has to increase very slowly with the population P_c. In addition, examination of density variation within several tracted cities confirmed directly the tendency to the law of exponential decrease outward. That law independently has been

[1] John Q. Stewart, "Suggested Principles of Social Physics," Science, 106, 1947, pp. 179-180.

published by the economist, Colin Clark.[2] See also a note by Stewart.[3]

Table 2 gives further confirmation and clarification of the pattern, showing how the central density, D_o, depended on the population, P_c, in 1940. These data suggest that the halving distance, b, increases from about a third of a mile for the smallest cities, of 2500 population, to perhaps 4 miles for the largest ones. Note that D_o is roughly proportional to the square root of P_c. As population increases the inc eased "attraction" (demographic gravitation) between people weighs down on and compresses the central population.

Evidence for the proportionality of the area, A, to the three-fourths power of P_c for cities in the United States and Europe was found to extend back at least to 1890, and there can be little doubt that cities were so structured a long time ago. Table 3 and Figure 1, compiled recently by Warntz, exhibit the same rule for British cities in 1951. Table 4 shows that cities in areas where general potential of population[4] is low tend to have larger areas than others of the same size-class located where the base potential (i.e., potential in the adjoining rural district produced by the remainder of the country) is high. This is confirmed also in the British study.

Full data as to the effect of base potencial of population in changing D_o, D', and b have not been compiled. It must be emphasized that the center of population density is not the business center except in the smallest cities. Of interest is the slow increase in the number of stories in dwellings in the central region of peak density from the smallest to the largest cities. Their height rises from one or two stories in cities of a few thousand people to six or seven stories in walk-up tenements in New York's lower East Side.

To explain the sharp edge density of a city -- a fall in a few hundred feet from 2000 people per square mile to rural densities of less than 200 -- one resorts to the concept of "cohesion" among city dwellers. The provision of the urban facilities which convert farm acreage into city lots requires joint cooperation supported by the inhabitants. Similarly the molecules within a water-drop stick together, and their mutual cohesional energy supports the "surface tension" of their boundary. Molecules cannot freely "evaporate" across the boundary -- unless by addition of heat the temperature is raised to the boiling point, when cohesion is overcome and the surface tension and the boundary itself disappear. Molecules in a gas have independent careers, and do not so stick together.

Over and above the capital energy released in any region, rural or urban, by demographic gravitation, additional energy is released from co-

[2]Colin Clark, "Urban Population Densities," Journal of the Royal Statistical Society Ser. A, Vol. 114, Part 4, 1951, pp. 490-496.

[3]John Q. Stewart, "Urban Population Densities," The Geographical Review, XLIII, 1953, p. 575.

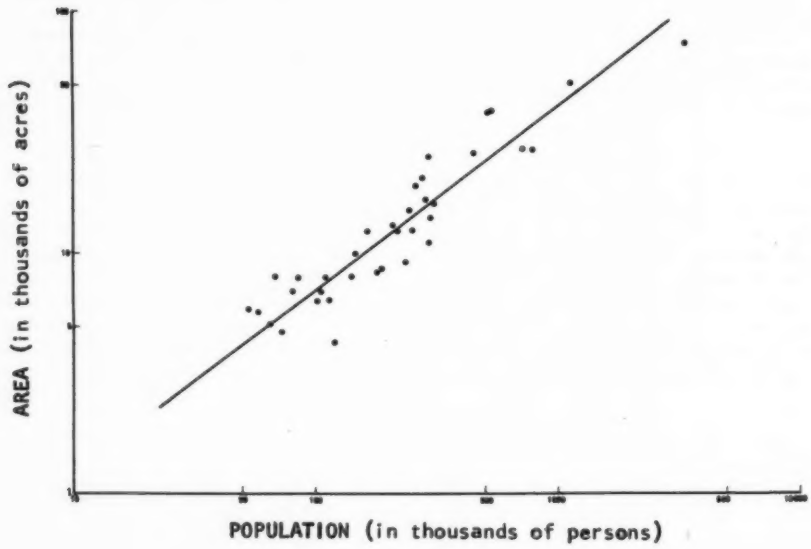
[4]For a recent discussion of demographic gravitation and potentials of population see John Q. Stewart and William Warntz, "Macrogeography and Social Science," The Geographical Review, XLVIII, 1958, pp. 167-184. See also Section 4 of the present paper.

TABLE 3. Areas and Populations of Cities in England and Wales, 1951.

RANK	CITY	POPULATION (in thousands)	AREA (in thousands of acres) ACTUAL	"EXPECTED"
1	London	3,348	74.8	105
2	Birmingham	1,112	51.1	46
3	Liverpool	790	27.3	35
4	Manchester	703	27.3	32
5	Sheffield	513	39.6	26
6	Leeds	505	38.3	25
7	Bristol	442	26.4	23
8	Nottingham	306	16.2	17
9	Kingston upon Hull	299	14.1	16
10	Bradford	292	25.5	16
11	Newcastle upon Tyne	292	11.1	16
12	Leicester	285	17.0	16
13	Stoke on Trent	275	21.2	15
14	Coventry	258	19.1	15
15	Croydon	250	12.7	14
16	Cardiff	244	15.1	14
17	Portsmouth	233	9.2	14
18	Harrow	219	12.6	13
19	Plymouth	209	13.1	13
20	Ealing	187	8.8	12
21-25	-	180	8.4	11
26-30	-	163	12.5	10
31-35	-	147	10.3	9.5
36-40	-	141	8.1	9
41-45	-	121	4.3	8
46-50	-	115	6.6	8
51-55	-	110	8.0	7.5
56-60	-	106	7.0	7.5
61-70	-	103	6.5	7
71-80	-	85	8.1	6.5
81-90	-	81	7.0	6
91-100	-	73	4.7	5.5
101-110	-	68	8.2	5
111-125	-	66	5.2	5
126-140	-	58	5.8	4.5
141-157	-	53	6.0	4.5

In Figure 1 land area in acres has been plotted against population on log-log paper for the 157 English and Welsh cities of over 50,000 persons in 1951. Land areas are in statute acres for the official political units (including inland water) as reported in the General Register Office's **Census 1951, England and Wales, Preliminary Report**, London, 1951. The leading twenty cities are shown individually. The next forty cities are indicated by means of medians of area and population for groups of five each. Groups of ten are employed for the next fifty cities. Following are two groups of

FIGURE 1. Relation of Area to Population for English and Welsh Cities, 1951.



fifteen each and finally one of seventeen. The above line indicates that the area of a city varies directly as the three-fourths power of the population. The equation relating area and population and used to compute "expected" values in Table 3 was found by a linear least squares solution of the logarithmic values of the variables for the first twenty cities. Stated in the power form this equation is: Area (in acres) equals 1.33 times population to the three-fourths power. If areas are to be stated in square miles, then the formula becomes: Area equals population to the three-fourths power divided by 481. (The coefficient of correlation, r, was found to be 0.87.) Compare this value of 481 to the ones similarly obtained in the United States of 357 in 1940 and 400 in 1890. In each of these cases the exponent of population was found to be three-fourths. Grouped data for the remaining 137 cities were also plotted in Figure 1 and indicate that the same relationship holds. The strong and well defined relationship of area to population is demonstrated.

Such scatter around the lines as exists can be further reduced by recourse to the macroscopic variable, base potential of population. Examination of the data clearly reveals that high potentials tend to constrict the area whereas at low potentials cities of a given population tend to occupy a larger area than indicated by the relationship of population alone. The effect of base potential is more pronounced for the less populous cities. (Potentials for the United Kingdom are mapped in Figure 2). In the final analysis, each city is a unique case and a complete explanation would entail a complex of microscopic factors such as local topography, a multitude of personal decisions, and the like. The point to be stressed is that social scientists must learn the value of first approximations and develop the abilities to make them. Of course the local planner whose job is an engineering one must learn to appreciate both the general and the particular.

TABLE 4. Area of Cities: Variation with Base Potential and Size.

Population	Potentials of population				
	170000	260000	370000	520000	730000
1,000,000+	--	--	172	127	299
500,000	--	--	61	46	--
200,000	45	53	38	18	19
100,000	50	38	20	19	10
50,000	13	14	10	10	5
33,000	11	12	9	7	6
25,000	7	9	6	6	4

All the 412 United States cities over 25,000 population in 1940 are represented. Their areas in square miles are presented in terms of median areas for groups of cities classified by population and by "base potential of population." The latter is the potential of population produced in the neighborhood of a city by all other people in the country, excluding those in the city itself. The rows refer to cities of the indicated population classes: namely, over 1,000,000, 500,000 to 1,000,000,.....25,000 to 33,000. The columns run according to base potentials, with median values 170,000, 260,000, 370,000, 520,000, 730,000 -- in 1940.

Evidently deviations in city areas from the one-variable relation of Table 1 are in part systematic, being well correlated with base potential as a second variable. Higher values of the latter are associated with a "compression" in city areas (related no doubt to greater land values of surrounding rural acreage). Warntz's examination of a number of British cities indicates the same effect (Table 3).

hesion in a city. This shows itself in increased activities of various sorts -- more local motor traffic, more local telephone calls. In large cities demographic gravitation is relatively more important, and cohesion in small ones.

We can also speak of "adhesion," a clinging of people to desirable land (or, negatively, their repulsion from unfavorable sites). A main transportation route running out of a city is an example of desirable land, and despite the gravitational attraction of the city, residences "rise" along it as water rises in a capillary tube (because of adhesion to the glass). But glass repels mercury (relative to the force of cohesion of liquid mercury particles for one another), so a surface of mercury is depressed within an inserted capillary tube. Similarly, people tend to avoid areas naturally or artificially "unattractive."

Our lake-front cities are not circular but are drawn out as semi-ellipses as the result of adhesion to the shore. Population density in the Pine Barrens of New Jersey has been below that in neighboring fertile land in every census from 1790 on. A small city in a narrow valley is elongated -- although the demographic gravitation usually is strong enough to crowd houses on steep slopes near the center.

While we have no adequate explanation as yet for the causes of the standard city pattern, cohesion must play a major role, along with demographic gravitation -- and adhesion can account for some of the observed distortions.

The daily, even hourly, movings about of people increase with techno-

logical and economic advancement -- just as the kinetic energy of molecules in the physical world increases with temperature. One's daily observations show how the automobile (much more effectively than the older forms of transport) operates to lower the boundary tension of cities. Widespread electric power networks, consolidated rural schools, state police protection of rural residents, and so on, have the same effect. Very important also is the fan of good highways extending out from every city, built and maintained at general expense rather than by direct assessment of those who choose to live along them. The combination of increased real incomes per capita with state and federal subsides for local improvements and for rural amenities is breaking down the old constant high edge density. In "star" patterns along the highways each city now in effect extends out well beyond the old bounding radius that was set by the high **D'**. The edge density now has perhaps fallen until it is almost indistinguishable from the rural density of the general countryside.

Plausibly, the standard city pattern is not otherwise changed. But many millions of people are listed now in the census as rural non-farm dwellers who might more properly be listed as urban. If this is so, the 1960 count of the urban fraction will, by former standards, be low by as much as 10 per cent.

Already in 1950, when the total population of the United States was 150,700,000, and there were 4284 cities, empirical regularities which had held from the first census[5] would have suggested an urban fraction of 64 per cent, instead of the tabulated 58.8. The reduction in the **edge density** of cities has changed the old balance of city and country. Further investigation is needed to arrive at a full theoretical explanation of the intra-urban equilibrium -- the density pattern outlined above and the changing rural-urban equilibrium. "Microscopic" studies of cities one at a time emphasize local distortions and necessarily have failed to reveal the existence of the standard city pattern.[6]

"Natural law" has been demonstrated to hold in social science as in physical science. City and regional planners who ignore tendencies of mass behavior do so at peril of serious failure in their designs.

2. Remarks on Other Characteristics of Cities

The residential city pattern has impressed itself also upon urban statistics other than those for population distribution. A correlation of urban rents with population and base potential existed in 1940 (before rent controls were effective as another influence).[7] Table 5, having the same sort of arrangement as Table 4, reveals the same two-variable dependence for municipal taxes per capita as for areas upon population and base potential -- although where high values of the latter reduce the area they

[5] John Q. Stewart, "Empirical Mathematical Rules Concerning the Distribution and Equilibrium of Population," **The Geographical Review**, XXXVII, 1947, pp. 461-485.

[6] Homer Hoyt, **The Structure and Growth of Residential Neighborhoods in American Cities**, Federal Housing Administration, Washington, D.C., 1939.

[7] John Q. Stewart, Chapter 2, **Theory in Marketing**, edited by Reavis Cox and Wroe Alderson, Chicago, 1950.

TABLE 5. Municipal Taxes per Capita: By City Size-Class and Base Potential.

Population	Potentials of Population				
	170,000	260,000	370,000	520,000	730,000
1,000,000+	--	--	36	38	73
500,000	--	--	37	72	--
200,000	19	29	18	54	70
100,000	19	16	16	41	52
50,000	17	15	15	57	50
33,000	14	22	14	37	51
25,000	14	19	14	38	39

The form is identical with that of Table 4, but with medians of municipal taxes, in dollars per year per capita in 1940, presented for the same 412 United States cities. Whereas higher base potential compressed the areas, it increased the city taxes.

TABLE 6. Excess of Urban Births over Deaths: By City Size-Class and Base Potential.

Population	Potentials of Population				
	170,000	260,000	370,000	500,000	730,000
1,000,000+	--	--	6.2	2.4	3.5
500,000	--	--	4.6	2.8	--
200,000	9.2	5.8	4.3	3.1	3.4
100,000	6.0	7.1	6.4	2.5	4.2
50,000	8.4	4.9	6.0	3.2	3.2
33,000	11.1	4.8	6.7	4.9	5.0
25,000	9.7	7.6	5.8	5.5	4.9

Some 412 cities as in Tables 4 and 5, 1940. Listed are medians of the excess of urban births over deaths (per 1000 total population annually). Here base potential was even more effective than city size in reducing the excess.

tend to increase the tax. Table 6 shows a like dependence in the excess of urban births over deaths per 1000 people.

Thus it is clear that the customary grouping of city statistics by size-class alone, regardless of base potential, can obscure significant regularities. Base potential is a macroscopic variable, which no amount of microscopic study can reveal.

The density pattern discussed in the first section is for the residential city, as has been stated. Before the advent of modern rapid transit the "occupational city" -- the spatial distribution of people at their jobs within cities -- necessarily closely coincided with the residential city. Nowadays it seems that proper census statistics -- which should be taken on a sampling basis -- would define the occupational city, superimposed upon the residential one.

We may guess that the mathematical equations for the two would be the same, and that the two "cities" would differ only in the numerical values of the parameters, P_c, A, a, D_o, b, D'. (From discussion above it is clear that these six represent only two independent variables, P_c and D', -- providing we ignore the effect of base potential, which is a third independent variable.)

Observations as one travels about the country indicate that the same factors which have reduced edge density, D', for the residential city have likewise reduced it for the occupational city. We shall return to this point below.

Hitherto unpublished data compiled from the 1940 United States census by Catherine Kennelly relate to the distribution among leading cities of a few sample occupations. Just as the city residential populations conform to the well-known "rank-size" or Pareto rule,[8] so also do the special subpopulations of residents listed city by city in specified occupations.

The rank-size rule states that, R being the rank of a given city in the list, the population for it equals MR^{-n}, M being the population of the largest city, of rank 1, and n an exponent constant for the sequence of all cities. When P_c is thus ranked, n for United States cities has long been unity, with New York at rank 1. If now "authors" are counted, New York again in 1940 had the largest contingent, 2765 of them, and the exponent n remained unity. Los Angeles ranked 2, whereas in the total population list it was 5. Denver was 7 instead of 24; Cambridge, Mass., 26 instead of 78; and Buffalo 42 instead of 14.

For real estate agents and brokers, n was again unity. New York led with 10884; Miami ranked 17 instead of 48, Pittsburgh 22 instead of 10. For operatives making automobiles and auto equipment, n was roughly 3/2; Detroit ranked 1 instead of 4; Flint was 2 instead of 56, New York was 12. This sort of statistic is a useful addition to the methodology of classifying cities with respect to their assumed special roles -- a tricky business.[9]

[8] John Q. Stewart, The Geographical Review, loc. cit., p. 464.

[9] Cf. an excellent and broadly stimulating paper by Otis Dudley Duncan, "Population Distribution and Community Structure," Cold Spring Harbor Symposia on Quantitative Biology, XXII, 1957, pp. 357-371.

What would one mean by the population, P_e, of the occupational city? It would be the number of individuals whose places of employment lie within the boundary of a given city. Walter Isard emphasizes three factors which relate to the concentration or dispersal of places of employment.[10]

The first of these, **economy of scale**, is related in part to the technological factors in a given industry and to the limited divisibility of certain factors of production. For certain kinds of manufacturing, low per unit costs of output can be achieved when many units per time period are produced at one location by a single firm. Thus, mass production economies up to a certain level of output tend to concentrate a large output at a given geographical point.

In the second place, **economies of localization** may be experienced when several similar firms in an industry cluster around a given geographical point in close proximity to one another.

Thirdly, **urban economies** which depend upon city utilities and facilities, local labor supply, etc., have brought the plants into actual cities.

Today ample observable evidence makes it clear that the lowering of a city's edge density for industry is becoming even more pronounced than for residences. No large-scale "decentralization" is resulting. Our once tightly bound cities surround themselves with "evaporated" dwellings and plants -- but the escape is only from local urban cohesion and not from the major national demographic gravitation. Only if the level of real national income increases tremendously will the high-density concentrations of residence and industry in the existing "manufacturing belt" thin out by removals to outlying situations of low population potential. To maintain the level of sociological intensity when people move farther apart demands that genuinely new and adequately profitable resources be tapped thereby.[11] Even then an associated increase in the birthrate probably would tend to maintain existing peaks of potential.

3. Rural Density; Interpenetrating Regions

Figure 2 is a hitherto unpublished map of population potentials for the United Kingdom, 1951. It permits testing there the relation of rural population density to the potential of population (produced by all the people). Results are given in Figure 3 and Table 7. Once again the rural density, D_R (in persons per square mile) is found to vary as the square of the potential, V, (in persons per mile). We have

$$D_R = k\, V^2,$$

where k (with the units stated) is determined for England and Wales as roughly 5.55×10^{-10}

Table 8 compares with this other values of k found as multipliers in the same proportionality of D_R to V^2 in the United States at different

[10] Walter Isard, **Location and Space-Economy**, New York, 1956, p. 172.

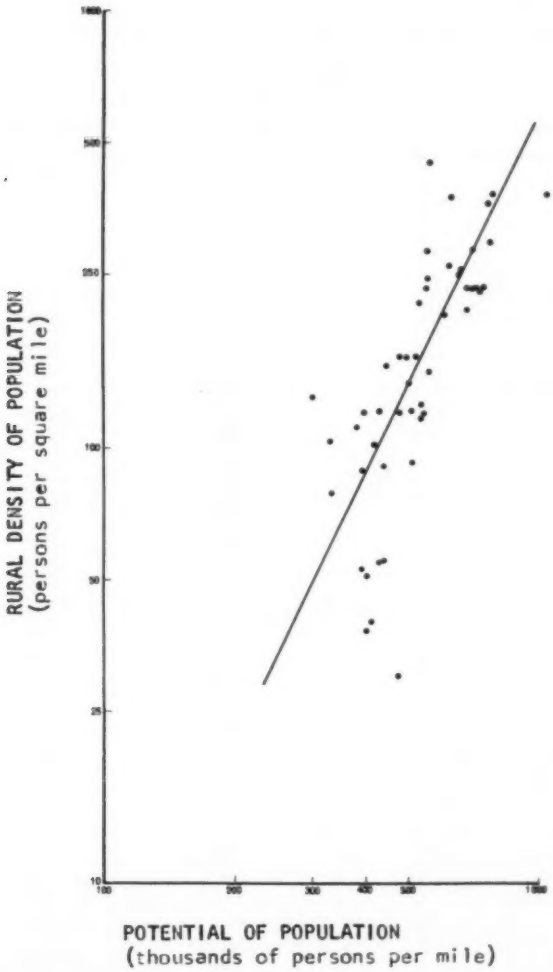
[11] John Q. Stewart and William Warntz, **The Geographical Review**, loc. cit., p. 173.

FIGURE 2. Potentials of Population in Thousands of Persons per Square Mile in the United Kingdom, 1951.

Contours of the "potentials of population" for the United Kingdom, 1951. Potential is a measure of the propinquity of people in the aggregate. Each individual contributes to the total potential at any place by an amount equal to the reciprocal of his distance away. This map of lines of equipotential of population in the United Kingdom, presented here for the first time, was drawn by means of logical contouring from values computed for 99 control points including the geographic centers of counties in England, Scotland, Wales, the Isle of Man, Northern Ireland, and the Irish Free State. The contours shown portray the base potentials from which each local city peak rises. The values of potential range from less than 100,000 persons per mile in the Shetland Islands to more than 1,900,000 persons per mile in the London Basin. If the influence at a distance of the rest of Europe's population were included in the computation of U. K. potentials, these potentials of course would be everywhere higher. The effect would be most pronounced in the southeast declining toward the

northwest. Examination of the general European map in John Q. Stewart, **Coast, Waves and Weather,** Boston, 1945, p. 166, shows that the general pattern of the contours over the United Kingdom would be little changed however. Special thanks are owed to David Neft, Christopher Warntz, and Francis Barkoczy for assistance in computing and drafting this map.

FIGURE 3. **Density of Population in Rural Districts and Potential of Population, English and Welsh Counties, 1951.**



The density of population in rural districts (averaged by counties) in England and Wales (in persons per square mile) for the 1951 census was found to be equal to 5.55×10^{-10} times potential of population (in persons per mile) to the second power. (The least squares coefficient of corre-

lation, r, is 0.70.) The same exponent of potential has been found in the observation of various other groupings of people. Therefore potential of population at a point as a macroscopic factor serves to provide a means of first approximation of the rural density over the area around that point. This estimate makes possible a better understanding of the significance of local or microscopic factors. (See also Table 7.)

TABLE 7. **Potentials of Population and Rural Population Densities, England and Wales, 1951.**

COUNTY	POTENTIAL OF POPULATION in thousands of persons per mile	RURAL DENSITIES in persons per square mile	
		ACTUAL	EXPECTED
Surrey	1,040	383	620
Derbyshire	786	386	350
Lancashire	780	298	345
Hertfordshire	775	364	340
Bedfordshire	745	235	315
Warwickshire	737	232	305
Nottinghamshire	723	234	295
Buckinghamshire	709	286	285
Staffordshire	703	234	280
Yorkshire (W. R.)	694	233	270
Essex	689	208	265
Leicestershire	665	257	250
Chesire	661	251	245
Durham	636	378	225
Oxfordshire	628	183	220
Kent	626	262	215
Worcestershire	604	202	205
Northamptonshire	573	146	180
Gamorganshire	569	453	180
Denbighshire	563	149	175
Gloucestershire	561	246	175
Berkshire	558	285	170
Sussex	557	233	170
Shropshire	550	120	165
Rutlandshire	542	126	160
Monmouthshire	541	119	160
Hampshire	536	215	160
Flintshire	533	332	155
Cambridgeshire	526	163	150
Herefordshire	516	93	145
Huntingdonshire	506	122	140
Suffolk	501	142	137
Wiltshire	499	163	135
Somersetshire	483	162	130
Lincolnshire	482	122	128
Radnorshire	481	30	127
Norfolk	450	155	111
Yorkshire (N. and E.R.)	449	91	110

Breconshire	443	55	107
Northumberland	440	54	106
Dorsetshire	432	122	103
Caernarvonshire	422	102	97
Merionethshire	412	40	92
Westmorland	404	51	89
Montgomeryshire	402	38	88
Anglesey	397	121	86
Cardiganshire	395	53	85
Cumberland	394	89	85
Carmarthenshire	381	112	79
Pembrokeshire	335	79	60
Devonshire	333	104	60
Cornwall	301	131	49

Presented here in tabular form are the data of Figure 3. Fifty-two counties in England and Wales are shown. (London and Middlesex counties had no rural population in 1951). The potential of population values (given above in thousands of persons per mile) are the ones computed as control point values for the drawing of the map in Figure 2. Actual density of population in rural districts (averaged by county) was obtained from the 1951 census data. The "expected" rural density was computed in the following way: Let D = rural density in persons per square mile and V = potential of population in persons per mile, then $D = 5.55 \times 10^{-10} V^2$.

Of course potential of population does not give final accuracy. It is, however, a single macroscopic integrative index which introduces a powerful unifying first approximation into the study of the geographical variation of sociological phenomena. Its effects must be supplemented by local factors in individual cases. The ruggedness of the terrain in such Welsh counties as Radnor, Montgomery, and Merion presumably offers a strong deterrent to rural settlement whereas the gentler slopes and more fertile soils in certain other places encourage it.

TABLE 8. Parameters for Rural Population Density.

Region	P_T millions	P_R millions	$10^{12}k$	w	q
U. S., 1900	76.0	45.8	880	0.603	0.111
1930	122.8	53.8	425	0.438	0.119
1940	131.7	57.2	351	0.435	0.106
Europe, 1930's	499.7	321.8	149	0.644	0.116
Mexico, 1930	16.4	11.0	3820	0.671	0.093
England & Wales, 1951	43.7	8.4	555	0.192	0.126

The columns list, respectively, for six different cases studied, total population, rural population, the value of k (multiplied by a million million), the numerical ratio, w, of rural population to total population, and, finally, another pure number, q, defined as equal to $P_T k/w$ (see text). In each case D_R, the rural density anywhere in the country, was approximately kV^2, V being the potential of population at the point, and k a parameter constant over the country at the time. The value of k as listed was determined by a median fit to the observed values of D_R, without any regard to the later computation of q. The greatest relative concentration of population near the center is indicated for Mexico, because q comes out smallest there.

dates, in Europe, and in Mexico, Respective total populations, P_T are listed, along with corresponding rural populations, P_R, and the ratio, w, of rural to total population. Tabulated also for each case is the computed value of a certain pure number, q, derived as follows:

Let A be the total area over which the populations P_T and P_R are dispersed. Therefore

$$P_R = \int D_R dA = k \int V^2 dA,$$

where k, being constant throughout the area, can be written outside the integral sign. The integration is over the entire area. Let there be a pure number q, where

$$q = \frac{P_T^2}{\int V^2 dA};$$

then

$$k = \frac{P_R}{P_T^2} \cdot q;$$

and, since

$$w = \frac{P_R}{P_T},$$

$$k = \frac{wq}{P_T} \quad \text{or} \quad q = \frac{P_T \cdot k}{w}.$$

Inasmuch as the potential V is that produced at any point by the whole population, P_T, divided by distance, while area always has the dimension of the square of distance, $\int V^2$ dA has the dimension P_T^2, and q comes out dimensionless, a pure number, as is w.

Table 8 shows that q is remarkably stable from case to case. For a population uniformly distributed over a circular disk (Table 10 below), q is readily computed as about 0.11.

The listed values of k in Table 8 were obtained by fitting a median straight line of slope 2, on log-log graph paper, to plotted values of observed D_R at various points of known potential, V. Then q was computed from the said k in each case. Now q could have been determined directly by summing terms V^2 dA.

It can be shown that $\int V^2$ dA computed over an assigned area for a fixed total population, P_T, is larger if the population is strongly concentrated near the center of the area -- and smaller if instead most of the people are placed near the boundary rim of the area. Hence a small value of q indicates central concentration of the population; a medium value of

TABLE 10. Potential within a Uniform Circular Disc, as a Function of the Distance from the Center.

Ratio c/r	Relative Potential
0.00	1.00
0.05	1.00
0.15	0.99
0.25	0.98
0.35	0.97
0.45	0.95
0.55	0.92
0.65	0.88
0.75	0.84
0.85	0.78
0.95	0.70
1.00	0.65

If a population, P, are distributed uniformly over a circular disk of radius r, the potential at the center is $2P/r$. This value is written in the table as, relatively, 1.00. The relative potential falls as shown with increasing distance c from the center, reaching only 0.65 at the edge (where c is equal to r). At a considerable distance beyond the edge, where the ratio c/r is large, the relative potential is approximately $r/2c$; thus at a distance of 5 radii from the center, where c/r is 5, it is 0.10, about. For at a distance the entire population P can be considered as though concentrated at the center, and the potential there is about P/c. Note that even at the very edge this approximate formula would give 0.50 for the relative potential, as compared with the accurate value of 0.65 tabulated.

q indicates roughly uniform distribution; and a larger q means boundary concentration.

So we have in q, thus computed, a new general index of distribution.

If, in computing V, people are assigned different weights it is required to take P_T not as the sum of the actual people, but as the sum of the product of people by their respective weights. None of the cases listed in Table 8 involved weights other than unity throughout. But weights suggested in different sections of the United States in 1940 would have increased the effective value of P_T by about 11,000,000 standard people.

The tendency of rural density to be proportional to the square of the potential, across a great country, is a well established fact. This indicates that such a country as the United States possesses demographic unity of a certain type. The applicability of the rank-size rule for cities also indicates a unity. Therefore we conclude that any subdivision into separated "regions" can have only special sorts of meaning -- there are no unique regions.

The fact that there are no unique regions has already been recognized in the wide use by regional analysts of two classes of regions: the "homogeneous" and the "nodal." As an example of the homogeneous, there are

regions where the rural density runs rather consistently high or low -- the Pine Barrens of New Jersey being an example of low density of population in every census from 1790 on. This is a kind of homogeneous region which is identifiable by its systematic deviation from a general regularity. Besides the proportionality of rural density to the square of the potential, other demographic or economic terms are known to vary as the square or as some other power of potential. For each such relation across the country, systematic deviations of actual from expected values may define a different set of regions.

Examples of nodal regions are the campus of a seasoned, privately endowed, "national" university and the dispersed residences of its students, a metropolitan newspaper and its territory, a seaport and its hinterland. They are all very special: the hinterland of a given port may differ for different commodities. The competition of nodal regions of a given sort may be "interpenetrating" as for colleges, or "all-or-none" as for newspapers in different great cities. Thus sales of the old St. Louis "Star Times" by counties around the city were proportional to the respective potentials of population there of the counties, but only out to where that newspaper's territory was rather sharply bounded by competition of papers in Kansas City, Chicago, Memphis. In contrast, leading colleges each draw students from every part of the country.

Even where subdivision into regions is for administrative purposes, as into states, counties, Federal Reserve districts, relations to population potential (or to income potential) make themselves evident, again confirming the general sociological unity of the country.[12]

4. Some Fundamentals of the Gravity Model

The recent rapid increase in the number of papers and articles employing the so-called "gravity models," while attesting (however belatedly) to the fact that distance is truly a dimension of social systems, has unfortunately brought with it much confusion concerning the exponent of distance to be employed. Whereas the "weights" assigned to people must be adjusted to fit observations, the function of distance is not such an arbitrarily adjustable parameter.

It appears that at present, potential of population, **V**, is the most widely known aspect of the gravity model. Knowledge of numbers of people, **P**, and their distances apart, **r**, are at once necessary and sufficient for the computation of potentials and the ultimate plotting of lines of equipotential on a map. Other demographic measures which can also be derived from these primitive quantities or "dimensions," and which are consistent with potential of population, include density, **D**, energy, **E**, and gradient, **g**. Time and velocity and acceleration are also consistent with the above but will not be considered here. Time is another primitive.

Potential of population is a "scalar" quantity (having no direction in space) and is equal to number of people divided by their distance away. The potential which a given population concentration creates at a distant point **Q** is:

[12] John Q. Stewart and William Warntz, **The Geographical Review**, **loc. cit.**, p. 173.

$$V_Q = \frac{P}{r} .$$

But, the total potential at any point is produced by all groups of people. If the distribution may be regarded as continuous over a surface, the following formula applies:

$$V_Q = \int \frac{1}{r} D \, dA.$$

where dA signifies an infinitesimal element of the area over which the integration is extended.

Since density of population has the dimensions of persons per unit area, and area has the dimension of distance squared, the derived potential of population has the dimensions of number of people per unit distance.

If the values of potential have been computed for a sufficiently large number of appropriately spaced "control points" then a map such as Figure 2 can be drawn to exhibit lines of equipotential or "contours."

If a constant contour interval is maintained throughout the map, the gradient of potential is inversely as the spacing of the contours -- gradient being the rate of change of potential with distance. Whereas potential is a scalar quantity, gradient is a vector directed at right angles to the equipotential contour at any place. Thus the unit of gradient has the dimensions of persons per mile squared. It should be noted that gradient and density thus have the same dimensions.

Although much analysis has centered upon the relation of the potential field to the geographical variation in certain sociological and economic phenomena, gradient has been shown to be important with regard to location theory.[13]

If only two separated groups of people are considered, their mutual "demographic energy" can be computed as:

$$E = \frac{P_1 P_2}{r}$$

The unit of energy therefore is persons squared per mile. The energy alternatively can be interpreted as the product of the population of either group times the potential contributed there by the other group. If one wishes to compute the energy of any group of people, say a city population, in relation to the total population, this can be found by multiplying the population of the city by the total potential contributed there by all people. Again the unit of energy is persons squared per mile.

Thus a rigorous and consistent set of measures exists of a population as it is distributed spatially. It should in particular be noted that the above formulas include without disagreement the case of W. J. Reilly's

[13] E.g., John Q. Stewart and William Warntz, The Geographical Review, loc. cit., p. 178.

"breakpoint" between competing market cities.[14] He published the rule of inverse square of distance because he investigated gradients, not potentials.

Both Reilly and Stewart found their respective exponents as empirical regularities without any a priori postulates.

5. Remarks and Formulas in Relation to Potential Theory

Mathematical statisticians have seldom blazed the difficult trail which leads from observations of the world to new branches of science. Their convention-bound methods come into play only after discovery of the leading concepts and relations of the new field. Statistics necessarily is microscopic. It deals with the data of special cases, and can work with broad data only if the cases already have been put together by connections discovered in studies in the field in question. The operation of a principle may be obscured by the simultaneous effects of other principles, and mere statistics cannot, in complete ignorance of conditions, sort out the principles.

Very wide ramifications are possessed by the formula for the influence of people at a distance -- as directly proportional to the number of people (weighted in whatever numerical ratio may be necessary) and inversely to the first power of the distance. The present paper indicates many of the ramifications, and in addition suggests some of the other factors which can add their own peculiar effects.

Naturally a mathematics-statistics hopper, however recondite and busy, which lumps confusedly together all these factors will be incapable of closely verifying the law of inverse distance. This description, we suggest, applies to the recent study by Hammer and Ikle of long-distance telephone calls and airplane trips between certain cities.[15] When in their study people were weighted equally, the formula for "mutual energy" of the two city populations (their product divided by the distance apart) was roughly verified. But, when the machine was allowed (subject only to the hidden conventions of mathematical statistics) to select the weights assigned to the people of each city, the exponent of distance came out as something like the negative three-halves power.

No comparative test was made with the approximate weights previously published by others -- namely, 2, 1, and 0.8 for different sections of the United States, nor, alternatively, per capita income. The actual observations of telephone calls and travel was not tabulated. One would not expect that airline trips among the given cities in the single month of March, 1950, would furnish a stable sample. Nor have experts of the A. T. and

[14] W. J. Reilly, "Method for the Study of Retail Relationships," University of Texas Bulletin, No. 2944, 1929.

[15] Carl Hammer and Fred Charles Ikle, "Intercity Telephone and Airborne Traffic Related to Distance and the 'Propensity to Interact,'" Sociometry, 20, pp. 306-316, 1957.

T. Co. compiled or released a broad table of phone calls among cities of all sizes. (Likewise the Federal Reserve System has made no adequate compilation of the inter-district flow of bank checks.)

It can be mathematically fallacious to assume a scalar "potential," or index, equal to population divided by distance to the nth power where **n** is any exponent. For example, if **n** is greater than 2, the index would have infinite values for finite distributions of population density. This is because $\pi r^2 D/r^n$ equals $\pi D/r^{n-2}$, and, if **n** is greater than 2, approaches infinity as **r** approaches zero. The same objection applies with the gradient if **n** is greater than 1 in the potential. Consequently it is a misnomer to use the name "gravity model" unless potential and gradient are defined as in the present paper. In three dimensional space there are not gravity models, only the gravity model. People are distributed in three dimensional space, although for convenience demography often can neglect the third dimension.

The published observational evidence for the usefulness of a scalar index operating to the inverse first power of distance is direct, varied and remarkably strong as social science evidence goes. Mathematical statistics can be a good servant; but only a weak social science, intent on becoming weaker, will welcome it as a master.

Local Peaks of Potential. -- Suppose a small city lies in a rural district where, apart from the city's own contribution, the contours of equipotential are, nearly enough, parallel straight lines, equally spaced. In section 4, gradient is defined as the rate of change of potential with the distance at right angles to the contours of equipotential. So here the general gradient, **g**, is uniform. The small city, if all alone, would be surrounded by contours more or less circular, sloped "downhill" radially outward. Superposition of these two sets of contours results in the single system illustrated schematically in Figure 4.

If instead of a city we had only a village, the hatched contour in the figure, and the circular ones included within its closed loop, would not come into existence. There would only be a short downhill ridge raised above the general level across the village. No local peak would then be formed, and no closed contours around the town. The condition that there be a local peak is roughly this: within the loop of the hatched contour the density of the local population must exceed the uniform gradient, **g**, -- the latter being measured at a distance away, outside the local influence. (As stated in section 4, gradient and population density have the same dimensions.)

The phenomenon of "urban sprawl" is well known near great cities. The gradient away from the central city is so large that even fairly sizable concentrations of local population do not match or exceed it. Local closed contours of potential are not achieved, and unless its internal cohesion is very great the would-be small city can have no independent demographic existence. As with the planet Saturn, the powerful gravitation of the central metropolis tears apart satellite towns. Is it possible for planners to provide the high cohesion which would preserve the character of such precariously perched neighborhoods?

A Theorem of Topology. -- The contours of potential are mathematically determined when the density of population is known throughout the whole

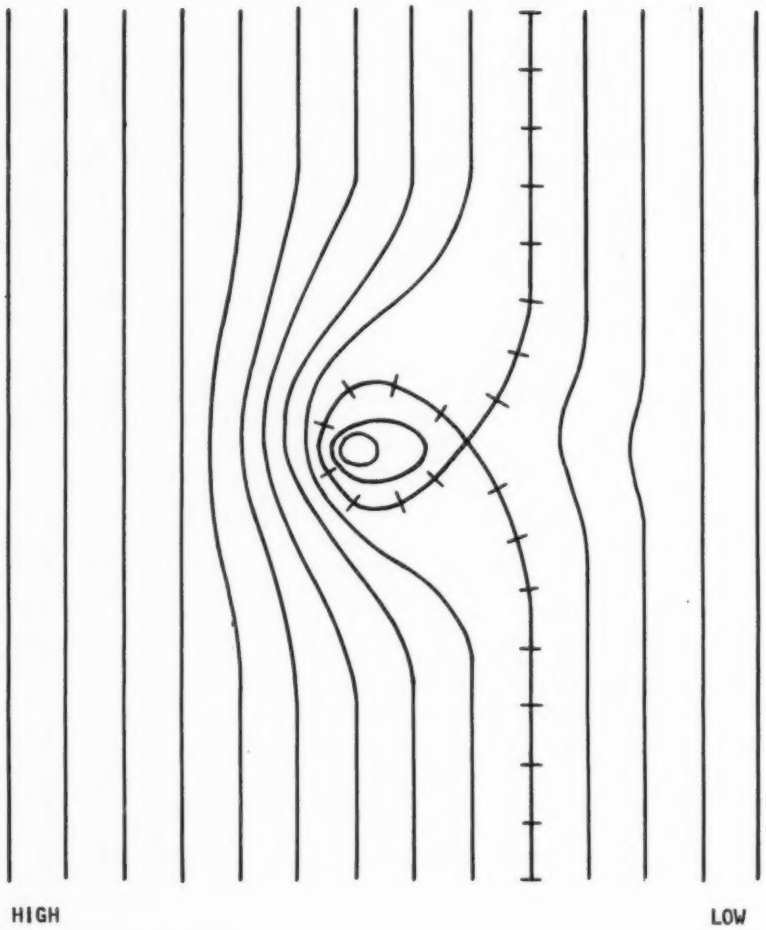


FIGURE 4. Schematic Representation of Contours of Potential near a Small City.

Far to the right the hatched contour would close around the main concentration of the country's population, forming thus the other, much larger, loop of a distorted figure eight. Its self-crossing marks a "saddle" or "pass." The "local peak" of the small city rises to the left, and the main massif of the nation to the right. On both sides of this axis the slope from the saddle is gently downhill. Farther to the left the city peak slopes sharply down. In general the slope or gradient at any point is perpendicular to the equipotential there, and, as this direction is ambiguous at the self-crossing, the gradient there is necessarily zero.

country. Certain distributions of population produce "pits" as well as "peaks" of potential. Of course the bottom of a pit may itself be at a high potential; the requirement merely is that the slope be upward in all directions in the immediate neighborhood.

An interesting theorem of topology applies within any closed contour of equipotential on a surface, as first stated scores of years ago by the great physicist, Clerk Maxwell, and later by the well-known mathematician, Marston Morse. The number of peaks plus the number of pits minus the number of passes equals one, always. Thus in the situation of Figure 4 a second and major peak is necessarily implied, far to the right (e.g., for the United States at New York City), and the portrayed parallel straight line contours really are parts of great sweeping arcs concave toward the major peak. In general, each separate city peak on the general massif is accompanied by its own neighboring pass on the New York side. Three otherwise isolated cities near together and forming a triangle would be separated by three passes on its three sides, and a pit would exist within the triangle.[15]

Potential Near the Center of a Small Populated Area. -- If a circular disk of radius r has uniform population density, $\bar{D}$, throughout, then the total population, P, is $\pi r^2 \bar{D}$. The increment of potential at the center produced by a ring of radius r and infinitesimal width dr is $2\pi \bar{D} dr$. Hence the total central potential is

$$V_c = 2\pi \bar{D} r = \frac{P}{(r/2)} .$$

For an irregular area, A, not too different in shape from a circle, the same formula approximately applies, if by r we understand the value determined by setting $\pi r^2 = A$, again, even though A be irregular. But if some actual area, A, is elongated, an ellipse will approximate it better. Mere inspection of the map will suggest a plausible approximation to α/β, the numerical ratio of the semi-major to the semi-minor axis. The area of the chosen ellipse would be $\pi \alpha/\beta$. However, Table 9 requires only estimation of the ratio α/β. One computes r as before, as though the assigned area A belonged to a circle. The potential near its center approximates

$$V_c = \frac{P}{fr/2} ,$$

where f is the number read from Table 9 as a function of α/β.

By these means it is easy to approximate the potential of a given population "on itself," inasmuch as if the area, A, is small the assumption of uniform density, $\bar{D}$, is nearly always moderately good.

Table 10 shows how the potential within the uniform circular disk falls off with increasing distance out from the center.

Additional Formulas for the "Standard City." -- In section 1 we had

[15] For four cities at the corners of a square, with a pit inside, see John Q. Stewart, **The Geographical Review**, loc. cit., p. 475.

TABLE 9. Table for Computing the Potential at the Center of an Elliptical Disk.

Ratio of axes α/β	Eccentricity	Value of f
1.0	0.000	1.00
1.5	0.745	1.01
2.0	0.866	1.02
2.5	0.917	1.06
3.0	0.943	1.09
4.0	0.968	1.12
5.0	0.980	1.15
6.0	0.986	1.20
8.0	0.992	1.27
10.0	0.995	1.33

For a uniformly distributed population over a circular disk (i.e., an ellipse with eccentricity zero), the potential at the center is equal to the population divided by half the radius, r. Values of f in the table (computed using elliptic integrals) permit ready computation of the potential at the center of an elliptical disk having a uniform distribution of population, as equal to the population divided by the quantity, half the equivalent radius times f. By equivalent radius we mean the radius of a circle which would have just the area of the chosen ellipse. The central potential of the elliptical disk is less than that of this circular disk (with the same population) in the ratio of 1 to f. This is because of the elongation of the ellipse, which puts the population farther away from the center. The first column gives the ratio of the semi-major axis, α, to the semi-minor one, β. The text describes the procedure when this table is used to approximate the potential "on itself" of a small populated area.

$$D = D_o\, 2^{-r/b},$$

which we now rewrite,

$$D = D_o\, e^{-r/b'},$$

where $b' = b/\log 2$, the logarithm being to base e.

The population of an infinitesimal ring of radius r and width dr is

$$dP_c = 2\pi D_o r\, e^{-r/b'} dr.$$

Hence the potential at the center is

$$V_o = 2\pi b' (D_o - D'),$$

D_o being as before the central density, and D' the edge density, with

$$D' = D_o\, e^{-a/b'}$$

Now

$$\int r\, e^{-r/b'} dr = -b'(r-b')e^{-r/b'},$$

so we see that the whole population out to the bounding radius a is

$$P_c = 2\pi \left[b'^2 D_o - b'(a + b')\, D' \right].$$

Empirically it was shown that the radius a increases in proportion to the three-eighth power of the population, P_c; and that D_o is proportional to the square root of P_c. The constants of the two porportions are invariant for all the cities at a given time (if we ignore the effect of base potential), but are expected to change from census to census, slowly. The edge density also was constant for the sequence of cities in 1940, ideally.

Hence the above equation determines b' as a function of P_c, at any rate for 1940 when we have all the empirical constants. Solutions can be tabulated; a/b' is a useful parameter.

World Map of Population Potential: -- Inasmuch as the inverse distance formula for potential is good in three dimensions, potentials at representative points all over the earth can be computed once population densities everywhere are known. Distances must be measured not as arcs on the surface of the globe but as straight-line chords direct from point to point. (For one thing, the arc distance is ambiguous because it can be measured either way around the great circle through any two points). Obviously populations of the various regions must be weighted. Per capita income weightings are perhaps available. "Social mass" per person would be better -- perhaps 3000 tons in the United States, only a few hundred pounds for Australian aborigines on their native deserts.

6. Sociological Intensity

In a quasi-equilibrium, the contours of population potential are observed in the simplest situations to outline coinciding "isotherms" of human social activity of various sorts. For instance, in the United States in 1950 the number of cities per 100,000 square miles of populations 25,000 to 35,000 varied about as the cube of the income potential.

However, other factors, as indicated in section 2, may combine with the potential to raise or depress social activity in a region. It is suggested that the indicated isotherms depict levels of "sociological intensity," and that this, rather than potential of population, is the active operating agent in bringing about such correlations as the one just stated for small cities. It corresponds to the statistical concept of temperature in physics.

While potential is objectively defined, the problem of an equally objective definition of sociological intensity or "social temperature," is still being studied. The relation for small cities refers to one type of combination of man-with-lands or income-with-land. Dr. J. D. Hamilton, one of our associates in the social physics group, has suggested a close relationship between rules governing such combinations and the rules of physical chemistry governing molecular compounds.

COMMUTING AND THE LABOR MARKET AREA[1]

by Roy Gerard

A definite geographic hierarchy of labor markets can be delineated from the small place up to and including the nation through the presence of one or more elements of homogeneity such as occupation, employer, residence, etc. Such a hierarchical setup at its apex includes many diverse occupations, employers and areas. The problem of delineation becomes less acute as the area under study is reduced in size. The practical geographic spread of the labor market is usually limited to the area where people work and live. This concept assumes specific geographic limits that can be related to the journey to work, more popularly known as the commuting range to an industrial or commercial center.

In most studies, this commuting range has been reduced to a concept involving a radius of a given number of miles from the central city. Or, a time consideration is introduced; the journey to work is said to be a radius of one hour travel time from the central point. Of course, the availability of good roads, nearby housing, educational facilities, etc. play important parts in setting limits for such areas.

Recent studies suggest that the concepts involving travel time, distance or even the accessibility of roads are too mechanical to define the labor market area accurately in terms of real commuting range. The availability of jobs has been introduced as a significant factor inducing commuting from home to a distant place of work. But few analyses point to the impressive role which availability of jobs as well as job security (also working and social arrangements) play in setting and maintaining a pattern of commuting.

Loss of jobs in home communities often leads workers to seek employment

[1] At the request of the editors, the material for this paper was developed from the writer's Ph.D. dissertation "The Syracuse Region from the Standpoint of Employment and Unemployment," (New School for Social Research, June 1957). Supporting statistics are also available in **Commuting in the Syracuse Labor Market Area from the Standpoint of Employment and Unemployment**, Business Research Center, Syracuse University, October 1957, mimeo.

in nearby or even distant cities.[2] If jobs are available in these other places, a pattern of commuting is built up which is difficult to disrupt. Where job security is achieved at plants in the nearby city, the worker tends to accept such a journey to work on a regular basis. He ordinarily keeps his place of residence since ties are strongest there and since distance to work is not unreasonable. The worker begins to question a return to work in his home community even when jobs there become available again. Job security, as exemplified in seniority, is cherished so much that the disadvantages of even a long journey to work are often overlooked.

A peculiar pattern of commuting is built up because of the job availability factor. The **theoretical** commuting range may extend **x** miles from the central city and potentially include all workers in that area. The **actual** commuting pattern, on the other hand, is sharply distorted as compared to the theoretical possibility.

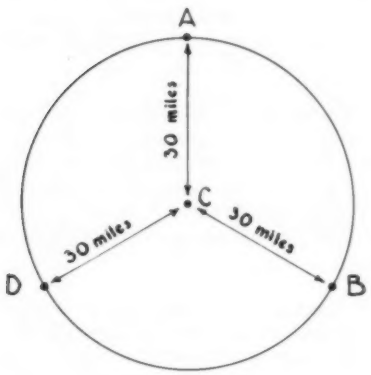


FIGURE 1a.

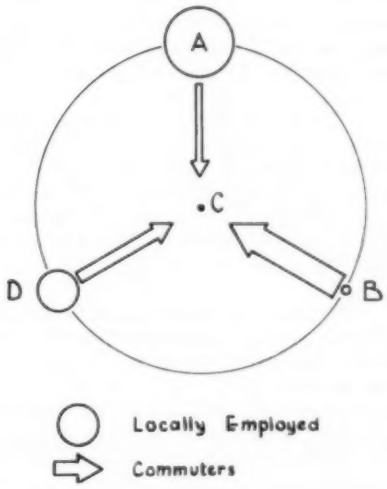


FIGURE 1b.

For example, assume 30 miles or a one-hour trip as the practical limits of a commuting area given good roads and equally favorable travel conditions from all points into the central city. Such may be represented by Figure 1a, where a symmetrical population distribution is assumed. However, workers in city **A** can find jobs in their hometown while workers in city **B** have difficulty in obtaining jobs locally and therefore take advantage of job opportunities open in central city **C**. Workers in city **D** find some jobs

[2]This is not always true. Workers in several upstate New York communities indicated that they would rather wait for job opportunities to develop in their locality than accept employment elsewhere, even if that employment is within commuting range. In one area, Auburn, New York, such waiting was rewarded when new job opportunities opened locally within a reasonable period after the worker had begun searching for employment.

available locally while others are open in the central city. The pattern of commuting, in terms of number of workers, rather than width of the commuting range, takes on a pattern depicted in Figure 1b.

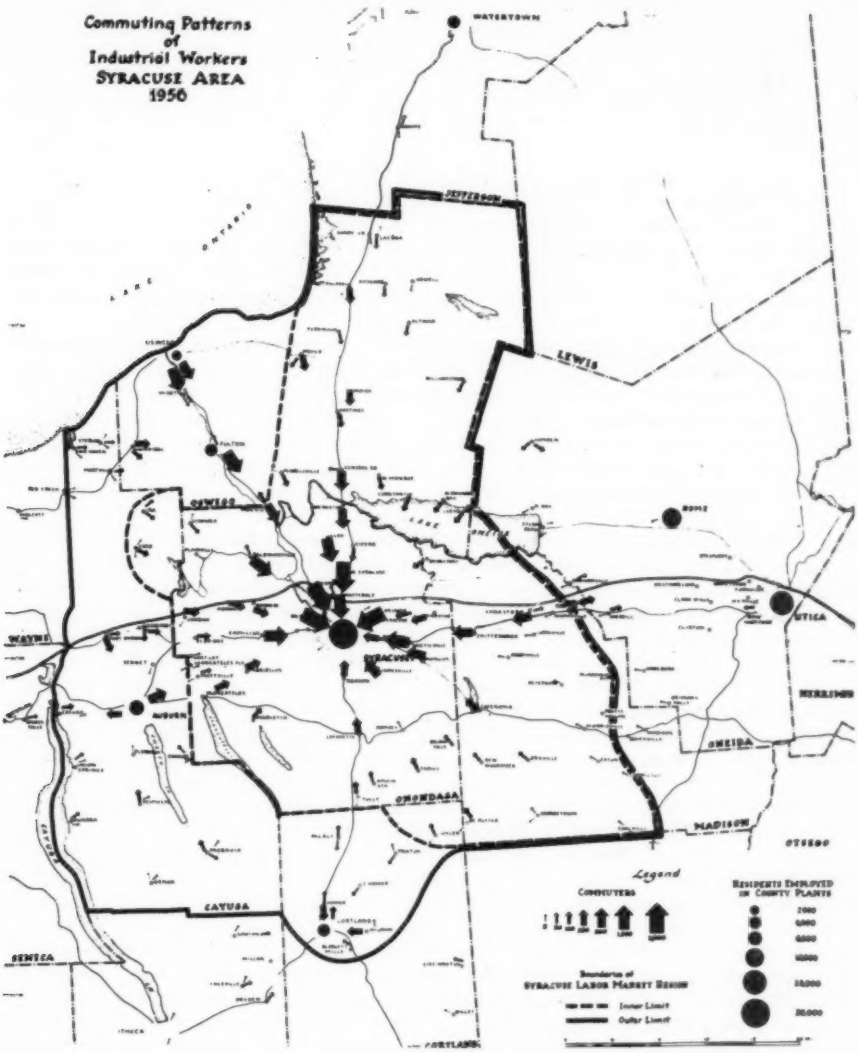
Such distributions have been most clearly shown in maps based upon data of the Federal Census of 1950 in West Germany. However, shadings were used to distinguish the volume of commuting between different geographic entities and the central city.

While the pattern of inflow from surrounding areas to a central city is obviously uneven, but still measurable, the pattern of outflow is generally far less significant. Only a trickle of workers commute from the central city area (which includes nearby industrial locations) to plants in distant localities. Usually, these workers are those employed in administrative or sales staffs of out-of-area firms who reside in the central city for cultural or other personal reasons. Factory or clerical workers are less affected by this type of "draw" of the central city. When jobs are no longer available locally (i.e. in the central city), these workers tend to **migrate** away from the central area, rather than live in the central area and **commute** to smaller areas.

This migration is induced not only by the workers' desire to live and work in or near a central city but also by hiring restrictions on the part of employers in smaller areas. These smaller area employers feel that workers from the central city would leave their jobs as soon as new job opportunities opened in the central city. Higher wages in the central city are a significant factor inducing the return of these workers from the smaller area to the central city.

Within the labor market, the pattern of commuting to the major center often obscures subsidiary commuting flows from one to another smaller area. These secondary flows build up in much the same way as the major flow into the central city. Size of city and distance are again less important factors than availability of job opportunities. There is, however, one major point of difference between the secondary and the primary commuting flows: the secondary flow involves such short distances and so little additional sacrifice by workers that it conceivably might fall outside the usual definition of commuting. Yet most workers consider that commuting is working anywhere but in their home community, no matter how short the distance travelled.

An example of the ebb and flow pattern in the central labor market and the subsidiary smaller labor markets can be illustrated by several specific examples. One such example is examined in the writer's study of the Syracuse (Central New York) area and commuting patterns of industrial workers. Map 1 presents the inflows and outflows for the labor market region dominated by Syracuse. There are noticeable deviations from the classic pattern. The size of commuting flow varies widely between places even where the distance and general accessibility (roads) are approximately the same. Thus, Auburn, New York (25 miles west of Syracuse) supplies far fewer commuters than Fulton (26 miles to the northwest) despite the fact that job opportunities are currently more plentiful in Fulton than in Auburn. Because of a coincidence in the past of job loss in Fulton and job availability in Syracuse, a pattern of commuting was built up that could not be disrupted even by the opening of new job opportunities in Fulton several years after the above pattern was established. This coincidence did not hold for Auburn



MAP 1.

workers. In the same category of coincidence but one that saw no later reopening of job opportunities locally is the commuting pattern of many Oswego (36 miles northwest) workers who took jobs in Syracuse. An equal number of other Oswego City residents filled job openings in Fulton that many Fultonians refused to accept having established seniority and work relationships in Syracuse.

Little or no flow takes place from Syracuse to the outside areas. Job opportunities locally as well as job restrictions by smaller area employers set an effective limit to any significant outflow by factory workers. As mentioned previously, executive and administrative staff are the only important commuters from the larger to the smaller area. And this group is negligible in size.

The labor market spreads most widely where job opportunities are in greatest abundance in a particular central area and when the labor supply in that area is least adequate to satisfy employer needs. In a period of full or near-full employment, such a condition occurs unevenly among labor markets. Some have a large local reservoir of labor to draw on while others, at the same time, find it necessary to range widely for workers.

The pattern of employment pulls from surrounding areas to a central city varies unevenly among the different smaller places. If job opportunities are available in a particular locality, workers who live there will tend to work there despite proximity to a much larger locality where job opportunities also exist. Workers in another locality more distant from the central city will take jobs in that city if few job opportunities exist in their area of residence.

The very process of growth of the local labor market -- particularly rapid during periods of large calls for labor -- leads to an equally sharp contraction during periods of retrenchment. The proportion of workers from outlying areas to the total central city work force is significantly larger for the more recently hired than for those workers with substantial seniority. As a result, cutbacks by central city factories at first affect outlying areas more than the central city. In fact, often a distorted picture of unemployment takes shape in an outlying area where local employment may be holding steady while unemployment induced by out-of-area layoffs builds up sharply.

The existence of seniority helps explain why workers commute to work and continue to do so despite the availability of job opportunities in their place of residence. Yet, on the other hand, this same factor is responsible for the early loss of jobs by those commuting workers who are the last hired with the least opportunity to build up seniority rights.[3]

Employment and unemployment in a region vary widely according to the tool of analysis used to define and demarcate the region. Thus, the usual work-residence study (in a period of full employment) of all workers in a group of plants in a central city will show on the average a narrow pattern of commuting because of the inclusion of many long-hired employees most of whom generally come from the local area. Another study, that of workers by recency of hire,[4] will show on the average a more extended region. The

[3]Another possibility encountered by the writer in interviewing plant personnel officials is discrimination against workers from distant areas leading to their layoff earlier than local workers despite greater seniority on the part of these commuters.

[4]See Herbert S. Parnes, A Study In the Dynamics of Local Labor Force Expansion, Ohio State University Research Foundation, 1951 (mimeo.).

layoff pattern, as indicated by filings for unemployment insurance benefits or by type of seniority list (90 days, 6 months, one year or longer), describes the widest region as well as the full effect of retrenchment by central city firms. In periods of substantial labor surplus, the work-residence analysis again depicts the smallest region since the most recently hired, generally the commuters, are no longer part of the central city plant work force.

National policy on labor markets is defined in terms of specific geographic areas, generally but not always limited to single counties. These areas are given special attention if a specified amount of unemployment in relation to the labor force can be shown to exist.[5] This method of defining an area has its shortcomings for, as shown above, few counties surrounding a central city can be conceived as being truly independent areas. While some local event may be a contributing cause of unemployment, generally the main factors of distress can be attributed to developments in a nearby central city which draws many of the smaller area's workers. The smaller area, dependent upon the larger area for jobs for many of its residents, "comes back" only with a return of prosperity in the larger area. Efforts to revive local industry or to bring in new industry often prove to be insubstantial compared with the effect of renewed commuting.

In an era of economic expansion by central city firms, the tendency of the labor market to widen results in a draw from smaller areas with a surplus of labor due to the weakness of local industry. This factor alone will reduce unemployment sharply. As previously indicated, by the same token, workers in the former surplus area have the least claim to seniority in plants that have expanded most recently. Hence, even a relatively small cutback in a central area factory usually affects these workers first. Being so dependent on the central area for employment for its workers, the smaller area slips into as well as out of its surplus classification freely. A redefinition of labor market areas would, therefore, appear to be in order based on the geographic spread of unemployment rather than on the static condition of long-established work-residence patterns.

Employment flow maps, such as the one in this study, have a specific value in describing main and subsidiary flows within a given area. The "inner" area of Map 1 (enclosed by dashed lines) may be considered as that region which includes all places where more residents commute to factories in the central city than to any other city. The "outer" area is that region in which central city influences are dominant over those of other major industrial centers. Two types of unemployment "dot" maps could conceivably be drawn: 1) a map of residences of laid-off workers according to seniority or recency of hire, 2) a map showing the number of unemployment insurance filings in nearby as well as central city offices as a result of layoffs by a central city firm or group of firms. These two types of maps would obviously show examples of regional structure different from each other and from Map 1. For instance, in a recent layoff by a large company in Syracuse, New York, (a central city), 27 per cent of the filings for unemployment insurance were by workers living in other than the county in which the layoff took place. Somewhat more than one-third of all workers with seni-

[5] Commuting patterns are taken into account only for purposes of determining whether a given county meets the criteria of a labor market area.

ority of 90 days or less resided outside the central county while only 14.5 per cent of the plant's entire work force were out-of-county residents.[6]

Briefly, in summary, labor market areas are defined by the level of prosperity or lack of prosperity in nearby central cities as well as by the physical fact of distance. The economic strength or weakness of labor supply areas is a major factor influencing their degree of dependency on the central area.

[6] It should be noted that this extreme variation of flow is far less significant for those plants with a stable work force and no recent expansion. In such instances, the layoff ratio of other areas to the central county would be close to that of the plant's basic work force.